

République Algérienne Démocratique et Populaire



Université des Sciences et de la Technologie
Houari Boumediene
U. S. T. H. B.

Faculté d'Electronique et d'Informatique
Département de Télécommunications

Thèse présentée pour l'obtention du titre de Doctorat d'Etat par

Salim Chitroub

En : Electronique
Spécialité : Electronique des Systèmes

Thème

Modélisation, Analyse, Classification et Fusion
d'Images de Télédétection Multisources Optiques et Radars

Soutenue le 09/02/2003 devant le jury composé de :

Mr M. AISSANI	Professeur	U. S. T. H. B.	Président
Mr A. GUESSOUM	Professeur	U. de Blida	Examineur
Mr A. KHELLADI	Professeur	U. S. T. H. B.	Examineur
Mr A. HOUACINE	Professeur	U. S. T. H. B.	Examineur
Mr Y. SMARA	Maître de conférences	U. S. T. H. B.	Examineur
Mr B. SANSAL	Professeur	U. S. T. H. B.	Rapporteur

Remerciements

*Avant tout, je remercie en premier lieu **Dieu, le Tout Puissant**, de m'avoir donné autant de courage, de volonté et de patience pour pouvoir achever ce travail. C'est grâce à **Dieu seul** que j'ai été recruté à l'Institut d'Electronique, inscrit en Thèse de Doctorat d'Etat et intégré un Laboratoire de recherche, au moment où tout cela était très difficile à avoir.*

J'exprime ma profonde gratitude à M. B. Sansal, Professeur à la Faculté d'Electronique et d'Informatique de l'U. S. T. H. B, pour m'avoir accordé toute sa confiance en acceptant de m'encadrer. Je tiens également à rendre hommage à sa rigueur et à sa disponibilité. Et pour finir, je ne cesserai de le remercier pour ses précieux conseils, qui ont été d'une grande utilité pour l'achèvement en bon terme de ce travail.

Je tiens aussi à remercier principalement Professeur M. Aissani, de la Faculté d'Electronique et d'Informatique, de m'avoir fait l'honneur d'accepter la présidence de mon jury et tout l'intérêt qu'il porte à mon travail.

Je remercie profondément M. A. Houacine, Professeur à la Faculté d'Electronique et d'Informatique de l'U. S. T. H. B, pour m'avoir offert un environnement de travail, pour ses conseils et pour l'aide qu'il m'a apportée.

Je remercie également M. A. Khelladi, Professeur à la Faculté de Mathématiques Pures et Appliquées de l'U. S. T. H. B, M. A. Guessoum, Professeur à Université de Blida, ainsi que M. Y. Smara, Maître de Conférences à la Faculté d'Electronique et d'Informatique de l'U. S. T. H. B, pour le grand honneur qu'ils me font en acceptant d'examiner mon travail.

J'exprime ma profonde reconnaissance à M. M. L. Tounsi, Enseignant-Chercheur à la Faculté d'Electronique et d'Informatique de l'U. S. T. H. B, pour sa précieuse aide et sa disponibilité.

Que tous ceux qui ont, de près ou de loin, apporté une aide à ce travail, trouvent ici l'expression de mes remerciements les plus sincères.

Sommaire

Introduction Générale	1
Chapitre 1 : Introduction au Traitement d'Images de Télédétection Spatiale Civile	
1. Introduction	5
2. Imagerie de Télédétection Spatiale Civile	6
2. 1. Formation des Images de Télédétection Optiques	6
2. 2. Formation des Images de Télédétection Radars	10
2. 3. Influence de l'Atmosphère	16
2. 4. Distorsions des Images de Télédétection	17
2. 4. 1. Distorsion Radiométrique	17
2. 4. 2. Distorsion Géométrique	18
2. 5. Contenu Informatif d'une Image de Télédétection	19
2. 5. 1. Aspect Spectral	19
2. 5. 2. Aspect Spatial	20
2. 5. 3. Aspect Radiométrique	20
2. 5. 4. Aspect Temporel	20
2. 6. Capteurs Imageurs en Télédétection Spatiale Civile	20
2. 6. 1. Imageurs Optiques	21
2. 6. 2. Imageurs Radars	22
2. 6. 3. Comparaison Entre Imageur Optique et Imageur Radar	23
2. 6. 4. Capteurs Imageurs Utilisés	23
2. 6. 4. 1. Série des satellites Landsat	23
2. 6. 4. 2. Série des satellites SPOT	24
2. 6. 4. 3. Instruments Végétation	25
2. 6. 4. 4. Système AirSAR	25
2. 6. 4. 5. Système SIR-C/X SAR	25
3. Traitement d'Images de Télédétection Spatiale Civile	26
3. 1. Modélisation d'Images de Télédétection	26
3. 2. Analyse Multidimensionnelle d'Images Optiques	28
3. 3. Analyse Multidimensionnelle d'Images Radars	31
3. 4. Classification d'Images de Télédétection	32

3. 4. 1. Classification ponctuelle	32
3. 4. 1. 1. Classification supervisée	32
3. 4. 1. 2. Classification non-supervisée	33
3. 4. 1. 2. a- Méthodes statistiques non supervisées	34
3. 4. 1. 2. b- Méthodes métriques non supervisées	34
3. 4. 1. 3 .Distinction entre méthodes statistiques et méthodes géométriques	35
3. 4. 2. Classification contextuelle	35
3. 4. 2. 1. Définition du contexte d'un pixel	35
3. 4. 2. 1. a- Contexte spectral/ polarimétrique	36
3. 4. 2. 1. b- Contexte spatial	36
3. 4. 2. 1. c- Contexte temporel	36
3. 4. 2. 2. Méthodes de classification contextuelle	36
3. 4. 2. 2. a- Méthodes structurales	37
3. 4. 2. 2. a- Méthodes statistiques	38
- Méthodes utilisant les transformations Markovienne d'image	38
- Méthodes utilisant la modélisation	40
3. 5. Fusion d'Images en Télédétection	41
3. 5. 1. Fusion bas niveau	43
3. 5. 2. Fusion de primitives	44
3. 5. 3. Fusion haut niveau	45
3. 5. 4. Fusion de classifieurs	46
3. 6. Filtrage, Détection de Contours et Segmentation d'Images	49
3. 6. 1. Filtrage d'images	50
3. 6. 2. Détection de contours	51
3. 6. 3. Segmentation d'image	51
4. Conclusion	52
Référence	53

Chapitre 2 : Caractérisation Statistique et Modélisation d'Images Optiques et Radars

1. Introduction	58
2. Caractérisation Statistique et Modélisation d'Images Optiques	59
2. 1. Modélisation par la Distribution Gaussienne	59
2. 1. 1. Sélection des échantillons d'apprentissage	60
2. 1. 2. Estimation du vecteur moyenne et de la matrice de covariance	61
2. 2. Cas d'Images Multispectrales de Dimension Elevée	63
2. 2. 1. Estimation du vecteur moyenne	63
2. 2. 2. Estimation de la matrice de covariance	64
2. 2. 2. a- Etat de l'art	64
2. 2. 2. b- Nouveaux estimateurs de covariance	68
b-1. Estimateur au sens du maximum de la moyenne des logarithmes des LOOL	68

- Sélection d'un mélange approprié	69
- Implémentation de l'estimateur	70
b-2. Estimateur au sens du maximum de la moyenne des logarithmes des BLOOL	73
b-3. Estimateurs de types LOOL mélangés	76
b-4. Estimateurs adaptatifs des covariances combinées	77
3. Caractérisation Statistique et Modélisation d'Images SAR	79
3. 1. Modélisation par la Distribution Complexe de Wishart	81
3. 1. 1. Formulation des images POLSAR	81
3. 1. 2. Caractérisation des cibles distribuées par la matrice de cohérence	82
3. 1. 3. Distribution de Wishart comme modèle pour les images POLSAR	84
3. 1. 4. Utilisation restrictive de la distribution de Wishart	86
3. 2. Modélisation par la Distribution K	87
4. Conclusion	89
Références	90

Chapitre 3 : Analyse Multidimensionnelle d'Images Optiques et Radars

1. Introduction	94
2. Analyse Multidimensionnelle d'Images Optiques	95
2. 1. Analyse en Composantes Principales : Approche Statistique	97
2. 1. 1. ACP standard appliquée en télédétection	97
2. 1. 2. Variantes de l'ACP appliquées en télédétection	100
2. 1. 2. a- ACP standardisée	101
2. 1. 2. b- ACP sélective	102
2. 1. 2. c- ACP à bruit ajusté	104
- Analyse théorique	104
- Extraction des images CP à bruit ajusté	106
- Implémentation	107
- Résultats pratiques	107
2. 1. 2. d- ACP à ajustement automatique du bruit	108
2. 2. Analyse en Composantes Principales : Approches Neuronales	110
2. 2. 1. Etat de l'art	111
2. 2. 2. Approche basée sur l'algorithme d'apprentissage de décorrélation	113
2. 2. 3. Approche basée sur l'algorithme d'apprentissage de réestimation	114
3. Analyse Multidimensionnelle d'Images POLSAR	114
3. 1. Problèmes Posés en Analyse d'Images POLSAR	115
3. 2. Etat de l'art	115
3. 3. Méthode pour l'Analyse Multidimensionnelle d'Images POLSAR	116
4. Conclusion	117
Références	117

Chapitre 4 : Classification Ponctuelle d'Images de Télédétection

1. Introduction	137
2. Classification Non-Supervisée	138
2. 1. Approches de Classification Non-Supervisée	138
2. 2. Avantages et Inconvénients de la Classification Non-Supervisée	141
3. Classification Supervisée	142
3. 1. Construction d'un Classifieur Supervisé	142
3. 1. 1. Classification par le maximum de vraisemblances	145
3. 1. 2. Classification par le minimum de distance de Mahalanobis	148
3. 1. 3. Classification par le minimum de distance Euclidienne	149
3. 2. Apprentissage dans la Classification Supervisée	149
3. 2. 1. Sélection des échantillons d'apprentissage	150
3. 2. 2. Phase d'apprentissage	151
3. 2. 3. Problématique de l'apprentissage supervisé	151
3. 2. 4. Comparaison entre les trois méthodes de classification	152
3. 3. Critères d'Evaluation de Classification	153
3. 4. Résultats expérimentaux	156
4. Classification d'Images Multispectrales de Dimension Elevée	157
4. 1. Approche Basée sur la Réduction de la Dimensionnalité	158
4. 2. Classification de Maximum de vraisemblances à Plusieurs Etapes	158
4. 3. Classification de Maximum de vraisemblances Rapide	159
4. 4. Classification de maximum de Vraisemblances Modifiée	159
4. 5. Approches Rapides de Classification	
de Minimum de Distances de Mahalanobis	159
4. 6. Classification de Minimum de Distances Euclidienne Modifiée	160
5. Conclusion	160
Références	161

Chapitre 5 : Classification Contextuelle d'Images de Télédétection

1. Introduction	167
2. Modélisation Markovienne d'Images	168
2. 1. Champs de Markov	169
2. 1. 1. Voisins, système de voisinage et clique	169
2. 1. 2. Définition du Champ de Markov	169
2. 1. 3. Distribution de Gibbs	169
2. 1. 4. Théorème d'Hammersly-Clifford	170
2. 1. 5. Topologie du système de voisinage	170
2. 2. Un Modèle Markovien Général d'Images	170
2. 2. 1. Estimateur Bayésien	171

2. 2. 2. Estimateur maximum a posteriori (MAP)	172
2. 2. 3. Estimateur modes a posteriori marginales (MPM)	172
2. 2. 4. Estimateur champ moyen "Mean Field (MF)"	173
2. 2. 5. La distribution a priori	173
2. 2. 6. Formalisme Bayésien-MRF	
pour la classification contextuelle d'images	173
2. 2. 7. La distribution a posteriori	174
2. 2. 8. Modélisation du terme de régularisation de la fonction d'énergie	176
2. 2. 8. 1. Modèle de Gauss	176
2. 2. 8. 2. Modèle d'Ising	176
2. 2. 8. 3. Modèle de Potts	176
2. 2. 8. 4. Première variante du modèle de Potts	177
2. 2. 8. 5. Deuxième variante du modèle de Potts	177
2. 2. 8. 6. Choix judicieux de la valeur de paramètre b	177
3. Optimisation	178
3. 1. Méthodes d'Optimisation Heuristiques et Métaheuristiques	179
3. 2. Optimisation Globale et Optimisation Locale	180
3. 3. Recuit Simulé	181
3. 3. 1. Dynamique de Metropolis	182
3. 3. 2. Modèle Mathématique du Recuit Simulé	183
3. 3. 3. Paramètres de Contrôle du Recuit Simulé	184
3. 3. 4. Implémentation du Recuit Simulé	185
3. 4. Recherche Tabou	186
3. 4. 1. Principe de Base	186
3. 4. 2. Critère d'Aspiration	188
3. 4. 3. Intensification	188
3. 4. 4. Diversification	189
3. 4. 5. Implémentation de la Méthode Recherche Tabou	189
3. 5. Algorithmes Génétiques	190
3. 5. 1. Principe de Base	191
3. 5. 2. Codage	192
3. 5. 2. 1. Codage binaire	193
3. 5. 2. 2. Codage réel	193
3. 5. 3. Population Initiale	193
3. 5. 4. Fonction d'Evaluation	193
3. 5. 5. Calcul de la Qualité	194
3. 5. 6. Opérateurs de Reproduction	194
3. 5. 6. 1. Sélection	194
3. 5. 6. 2. Croisement	195
3. 5. 6. 3. Mutation	196
3. 5. 7. Quelques Résultats Théoriques	198
3. 5. 7. 1. Principe des alphabets minimaux	198
3. 5. 7. 2. Principe de pertinence des schémas	198
3. 5. 8. Programmation Génétique	199
3. 5. 8. 1. Génotype et phénotype	199
3. 5. 8. 2. Fonctions de base	199
3. 5. 8. 3. Déroulement d'un programme génétique	199
3. 5. 8. 3. a- Fonction d'évaluation	200

3. 5. 8. 3. b- Reproduction	200
3. 5. 8. 3. c- Croisement	200
3. 5. 8. 3. d- Mutation	201
3. 5. 9. Classification Contextuelle Utilisant la Programmation Génétique	202
4. Résultats Expérimentaux	204
5. Conclusion	206
Références	207

Chapitre 6 : Fusion de Données en Télédétection

1. Introduction	213
2. Fusion d'Informations ou Combinaison d'Informations	214
3. Fusion-Classification d'Images Brutes	215
3. 1. Notion d'Incertitude	216
3. 2. Approches Type Fusion-Classification Fondées	217
4. Fusion de Classifieurs	219
4. 1. Méthodes Proposées : Fusion de Classifieurs d'Imagerie	220
4. 1. 1. Propriétés désirées pour une fusion optimale des classifieurs	220
4. 1. 2. Indépendance des classifieurs	221
4. 1. 2. 1. Dépendance et corrélation	221
4. 1. 2. 2. Indépendance statistique des classifieurs	222
4. 1. 2. 3. Indépendance conditionnelle des classifieurs	222
4. 1. 3. Hypothèses	223
4. 1. 4. Méthode Bayésienne de fusion des classifieurs	223
4. 1. 4. 1. Méthodologie	223
4. 1. 4. 2. Estimation des vraisemblances des classes	224
4. 1. 4. 3. Règles de décision après la fusion	226
4. 1. 5. Méthode de la théorie de l'évidence de fusion des classifieurs	226
4. 1. 5. 1. Première étape de fusion	227
4. 1. 5. 2. Deuxième étape de fusion	228
4. 1. 5. 3. Règle de décision. Extrait de naissance	229
4. 1. 6. Méthodes de vote majoritaire de fusion de classifieurs	229
4. 2. Résultats Expérimentaux	231
5. Conclusion	233
Références	234

Annexe : Publications

Introduction Générale

La télédétection spatiale permet de mesurer à distance, grâce à des capteurs embarqués sur les satellites, certaines propriétés radiatives d'un milieu naturel, caractérisées par la quantité d'énergie réfléchie par le sol en direction du capteur. Généralement, ces mesures sont traduites ou transformées sous forme d'images acquises dans différentes bandes spectrales du spectre électromagnétique à différentes résolutions spatiales et spectrales. Les nombreux capteurs disponibles permettent ainsi d'étudier de façon détaillée, précise et à moindre coût l'évolution des changements terrestres sur des surfaces étendues.

Les images fournies par les capteurs peuvent résulter d'un éclairage solaire, capteurs optiques, ou d'une illumination active de l'objet terrestre résultante d'un éclairage provenant d'une source émettrice, capteurs radars. Ces images permettent ainsi de fournir des informations sur l'objet terrestre grâce à l'interaction matière-rayonnement déterminée par les caractéristiques du rayonnement utilisé (longueur d'onde, état de polarisation, intensité, ...). Cela signifie, en particulier, que la caractérisation complète d'un objet terrestre ou d'un phénomène doit faire intervenir des informations différentes fournies par une gamme de capteurs couvrant les régions les plus significatives du spectre électromagnétique.

Par rapport aux instruments optiques, les systèmes radars hyperfréquences présentent deux avantages majeurs : tout d'abord, le radar est un système actif qui s'affranchit des variations du rayonnement solaire et garde l'avantage par sa capacité d'observation de jour comme de nuit ; ensuite, dans la gamme des fréquences utilisées, de 1 à 10GHz, les ondes électromagnétiques sont quasiment insensibles aux composantes de l'atmosphère. En fait, les inconvénients majeurs des capteurs optiques est qu'ils ne peuvent pas fournir des informations sur la surface imagée pendant la nuit ou bien si le ciel est nuageux. D'autant plus que l'atmosphère par ces phénomènes d'absorption et de diffusion, perturbe la réponse réfléchie par la surface. Cependant, contrairement à l'imagerie optique, le traitement et l'interprétation des images radars sont souvent difficiles car elle est influencée par plusieurs paramètres. La sensibilité de l'imagerie radar à la géométrie des cibles permet de fournir des images qui renseignent sur la rugosité, la topographie et l'humidité du sol, mais ceci est fortement dépendant du type de polarisation et de la

longueur d'onde du signal émis. L'image radar est sévèrement dégradée par le speckle. Le speckle est un phénomène qui se produit chaque fois qu'il y a des interférences entre les réponses réfléchies. Il a les caractéristiques d'un bruit multiplicatif qui nuit à la lisibilité des images. Il présente des défis pour un traitement efficace des images radars.

L'exploitation directe des images de télédétection est impossible, en raison des faiblesses inhérentes aux différents systèmes d'acquisition, optiques et radars, et la nature des images obtenues, ainsi que la quantité d'information acquise relative à une scène imagée, qui est considérable dans le cas des images optiques multispectrales et radars polarimétriques. La complexité des informations contenues dans ces images rend leur analyse par un photo-interprète lourde et ardue. Des traitements sont alors nécessaires pour exploiter ces images.

Les images multispectrales optiques de dimension élevée de la nouvelle génération des systèmes d'acquisition posent le problème des techniques et méthodes utilisées pour les traiter. Les méthodes et les techniques développées pour le traitement des images de dimension faible ne sont pas valables pour traiter les images de dimension élevée. Un autre problème posé est le temps de calcul qui augmente avec le nombre d'images spectrales traitées. Ainsi, afin d'exploiter la richesse d'informations contenue dans ce type d'images, de nouvelles méthodes et algorithmes doivent être développés.

Les images acquises sur une même scène par différents capteurs sont partiellement redondantes puisqu'elles représentent la même scène, et sont partiellement complémentaires, puisque les capteurs ont des caractéristiques différentes, et que les mécanismes d'interaction de l'onde électromagnétique et de la surface imagée diffèrent. Autrement dit, l'examen de chacune des images d'une même scène, produites par les différents capteurs, montre qu'il existe des informations significatives propres à chacun des capteurs qui peuvent être détectées facilement dans une image et difficilement dans une autre image. Il montre aussi qu'il y a une certaine relation existant entre les images des différents capteurs. En outre, l'information fournie par chaque capteur est incomplète et imprécise. Il est intéressant alors d'une part de traiter la redondance entre les différents capteurs pour réduire l'imprécision contenue dans les données de chaque capteur, et d'autre part d'utiliser leur complémentarité, en exploitant les informations significatives portées par chacun des capteurs, pour obtenir une interprétation qui ne pourrait être obtenue à partir des données d'un seul capteur. Ceci nécessite l'exploitation simultanée de l'ensemble des images fournies par les différents capteurs. Dans ce cas, le problème qui se pose réside dans la recherche d'une manière efficace pour l'exploitation simultanée de ces différentes sources d'informations. Il s'agit ici d'élaborer un processus appelé fusion de données multisources. La stratégie à suivre dans la fusion devrait dépendre de l'objectif visé et de la manière dont la fusion sera effectuée. Ainsi, une interprétation plus cohérente de la scène imagée, dans laquelle l'incertitude associée est diminuée, peut être obtenue.

Dans le cadre de cette thèse, le travail que nous avons réalisé a consisté à résoudre les problèmes de modélisation, d'analyse, de classification, et de fusion d'images de télédétection civile multisources, optiques et radars. Le but est d'apporter une contribution par l'élaboration de nouvelles méthodologies et de nouveaux algorithmes de traitement d'images de télédétection pouvant être utilisés dans n'importe quel domaine d'application. La thèse est organisée en six chapitres.

Le premier chapitre est consacré à familiariser le lecteur avec le domaine de traitement d'images de télédétection spatiale civile. Ainsi, le processus de formation des images de télédétection spatiale, optiques et radars, est présenté. L'influence de l'atmosphère et les différentes distorsions que ces images ont subies sont données. Le contenu informatif des images de télédétection est donné et les différents capteurs imageurs utilisés dans les applications civiles sont exposés. Les méthodes de traitement d'images utilisées en télédétection civile sont abordées.

Dans le deuxième chapitre, le travail qui a été réalisé consiste à modéliser les images, optiques et radars, par des modèles probabilistes appropriés et donner les méthodes d'estimation de leurs paramètres statistiques. Pour les images optiques, le modèle Gaussien est proposé et la problématique de l'estimation de ces paramètres statistiques est abordée. Le cas des images multispectrales de dimension élevée a été traité en mettant l'accent sur l'importance de l'exactitude de l'estimateur de la matrice de covariance de la classe. Ainsi, des méthodes proposées pour l'estimation de la matrice de covariance ont été étudiées tout en faisant ressortir les inconvénients et les avantages de chacune d'elles. Pour la modélisation des images radars, deux modèles ont été étudiés en détail : La distribution de Wishart et la distribution K. Nous avons montré que l'utilisation de la distribution de Wishart est restrictive dans la pratique pour des raisons qui ont été justifiées. Le développement théorique de la distribution K comme modèle pour les images radars ainsi que l'estimation de son paramètre a fait l'objet d'un article publié dans une revue internationale.

Le chapitre trois est consacré à l'analyse multidimensionnelle d'images multisources optiques et radars. Pour les images optiques, la méthode d'Analyse en Composantes Principales (ACP) et ses variantes appliquées aux images optiques de télédétection ont été développées. Une nouvelle approche neuronale de l'ACP des images optiques a été proposée et publiée dans une revue internationale. L'approche est basée sur l'algorithme d'apprentissage de décorrélation. Nous avons aussi proposé une nouvelle approche neuronale de l'ACP et ses variantes qui est basée sur l'algorithme d'apprentissage de réestimation. L'approche a fait l'objet d'un autre article soumis pour la publication dans une revue internationale. Des études comparatives entre les approches statistiques et les approches neuronales de l'ACP et ses variantes ont été effectuées.

Concernant l'analyse multidimensionnelle des images radars polarimétriques, la problématique du bruit multiplicatif provoqué par le phénomène speckle a été soulevée tout en mettant l'accent sur l'impossibilité d'appliquer les méthodes développées pour l'analyse des images optiques sur les images radars. Dans ce sens, une nouvelle méthode pour l'analyse des images radars polarimétriques a été proposée et publiée dans une revue internationale. Dans cette publication, la procédure d'implémentation de la méthode proposée est donnée en se basant sur un concept théorique prouvé. Un état d'art sur les différentes méthodes proposées dans la littérature pour l'analyse des images radars polarimétriques a été aussi donné dans le but de faire une étude comparative avec notre méthode proposée.

La classification des images multisources optiques et radars fait l'objet du chapitre 4. Tout d'abord les méthodes de classification ponctuelle des images de télédétection ont été développées. Une approche de validation de la précision de classification, sans avoir recours à la réalité de terrain, est aussi donnée. Les méthodes de classification contextuelle sont ensuite développées dans le chapitre 5. Ces méthodes sont développées afin de pallier aux inconvénients inhérents aux méthodes de classification ponctuelles. Ainsi, la modélisation Markovienne en traitement d'images a été abordée et des méthodes d'optimisation ont été développées.

La fusion de données en télédétection est l'objet du dernier chapitre. Le problème de définition exacte du terme de fusion de données est soulevé. Les ambiguïtés dans la signification ainsi que l'objectif inhérent à l'utilisation de la fusion de données sont mises en évidence. Ensuite, une méthode de fusion d'images brutes multibandes/ multisources, basée sur la théorie de l'évidence, est proposée. Un autre type de fusion d'informations est abordé : il s'agit de la fusion des classifieurs. Le but est d'améliorer la classification de la scène imagée, en considérant la fusion comme une deuxième classification qui succède à celle des classifieurs. Trois méthodes de fusion de classifieurs ont été proposées. La première méthode est basée sur le principe de la fusion Bayésienne. La deuxième méthode fait appel à la théorie de l'évidence dans son concept théorique. Quant à la troisième méthode, elle est basée sur le principe de vote majoritaire. Les trois méthodes cumulent les avantages de chaque classifieur sans en cumuler les inconvénients.

A la fin de cette thèse, nous avons inséré les textes intégraux des différentes publications parues les revues : *Signal Processing*, *Intelligent data Analysis*, *An International Journal*, *International Journal of Remote Sensing* et *SPIE*.

Chapitre 1

Introduction au Traitement d'Images de Télédétection Spatiale Civile

1. Introduction

La télédétection spatiale, ou l'observation de la Terre depuis l'espace, contribue, dans une large mesure, au suivi et à la compréhension de nombreux processus de l'environnement. Grâce aux images fournies par une large gamme de capteurs embarqués dans les satellites de télédétection civile, cette dernière est devenue une technique de plus en plus utilisée que ce soit pour la gestion de l'environnement, suivi et/ ou prévision des catastrophes naturelles, développement des modèles environnementaux globaux, localisation et suivi des ressources naturelles, étude de la diffusion de la maladie ou des parasites des récoltes, localisation de réseaux routiers, identification de failles sismiques, planification de grands projets de construction, etc.. [1]-[6].

Le présent chapitre est constitué de deux paragraphes essentiels. Le premier paragraphe est consacré à l'imagerie de télédétection spatiale civile. La formation des images de télédétection spatiale, optique et radar, est exposée. L'influence de l'atmosphère ainsi que les différentes distorsions, qui sont introduites lors de l'acquisition des images de télédétection, sont aussi données. Enfin, le contenu informatif d'une image de télédétection et les principales caractéristiques des différents capteurs imageurs utilisés sont présentés.

Dans le deuxième paragraphe, le traitement d'images de télédétection spatiale civile est abordé. Nous présentons seulement les méthodes de traitement d'images qui seront détaillées dans les chapitres suivants. Ceci consiste à donner un état de l'art des différentes étapes constituant notre méthodologie globale pour le traitement d'images de télédétection. Ainsi, nous évoquons la modélisation, l'analyse, la classification, et la fusion des images de télédétection spatiale civile. D'autres méthodes de traitement, qui ne font pas partie de notre méthodologie globale adoptée pour le traitement d'images de télédétection telles que : la détection de contours, le filtrage et la segmentation d'images, etc..., ne sont pas détaillées mais le lecteur intéressé peut se référer aux références citées dans la bibliographie.

2. Imagerie de Télédétection Spatiale Civile

Les images de télédétection spatiale civile sont obtenues à partir des mesures de rayonnements réfléchis ou diffractés par la surface terrestre. Ces mesures dépendent principalement de l'interaction du rayonnement électromagnétique avec l'objet terrestre observé ainsi que de la résolution spatiale du capteur de mesure et de la longueur d'onde correspondante à la zone du spectre électromagnétique dans laquelle le capteur est opérationnel (voir figure 1) [7]-[10]. Cependant, les images que l'on acquiert sont dégradées par l'atmosphère, le système optique/ ou radar et le capteur. Les signaux réfléchis par la scène observée traversent d'abord l'atmosphère où ils subissent les effets de la turbulence, puis le système optique où agissent principalement la diffraction et les aberrations. L'image d'intensité qui se forme ainsi au niveau du capteur est dégradée. Elle subit alors un échantillonnage, par intégration sur les éléments photosensibles (pixels) du capteur. L'intégration n'étant pas instantanée, l'image subit les effets du mouvement de l'instrument. Les signaux réfléchis sont ensuite convertis en courant électrique entachés de bruit de diverses origines. Ces courants subissent une quantification avant de former une image, c'est l'étape de mesure, qui contribue encore à la dégradation.

L'objectif des deux sous-paragraphes suivants consiste à décrire comment une image de télédétection optique/ ou radar se forme-t-elle, à bord d'un satellite d'observation terrestre, à partir de la scène que l'on observe. Tout d'abord, la scène observée est supposée se trouver à l'infini, car elle est observée à une distance qui est suffisamment grande devant sa taille. D'autre part, on suppose que toutes les corrections géométriques ont été effectuées, de manière à ce que les distorsions géométriques puissent être considérées comme négligeables. Cette hypothèse nous permet de modéliser la chaîne d'acquisition afin de ne prendre en compte que les effets de flou, de bruit et les défauts radiométriques. En fait, les différentes distorsions (radiométriques et géométriques) introduites lors de l'acquisition des images sont données en détail dans le sous paragraphe qui suit. Enfin, l'influence de l'atmosphère sur le choix judicieux des bandes spectrales est abordée et les principales caractéristiques des capteurs imageurs utilisés en télédétection spatiale sont données dans le dernier sous-paragraphe.

2. 1. Formation des Images de Télédétection Optiques

La chaîne de formation de l'image spectrale discrète (acquise par un capteur optique correspondant à une bande spectrale étroite de la partie optique du spectre électromagnétique) à partir de la scène observée et donnée dans la figure 2 est résumée par le tableau 1 [78], [80]. Notons que le satellite nous fournit une image multispectrale constituée de plusieurs images spectrales dont le nombre dépend du nombre de capteurs du système d'acquisition embarqués sur le satellite (voir figure 3).

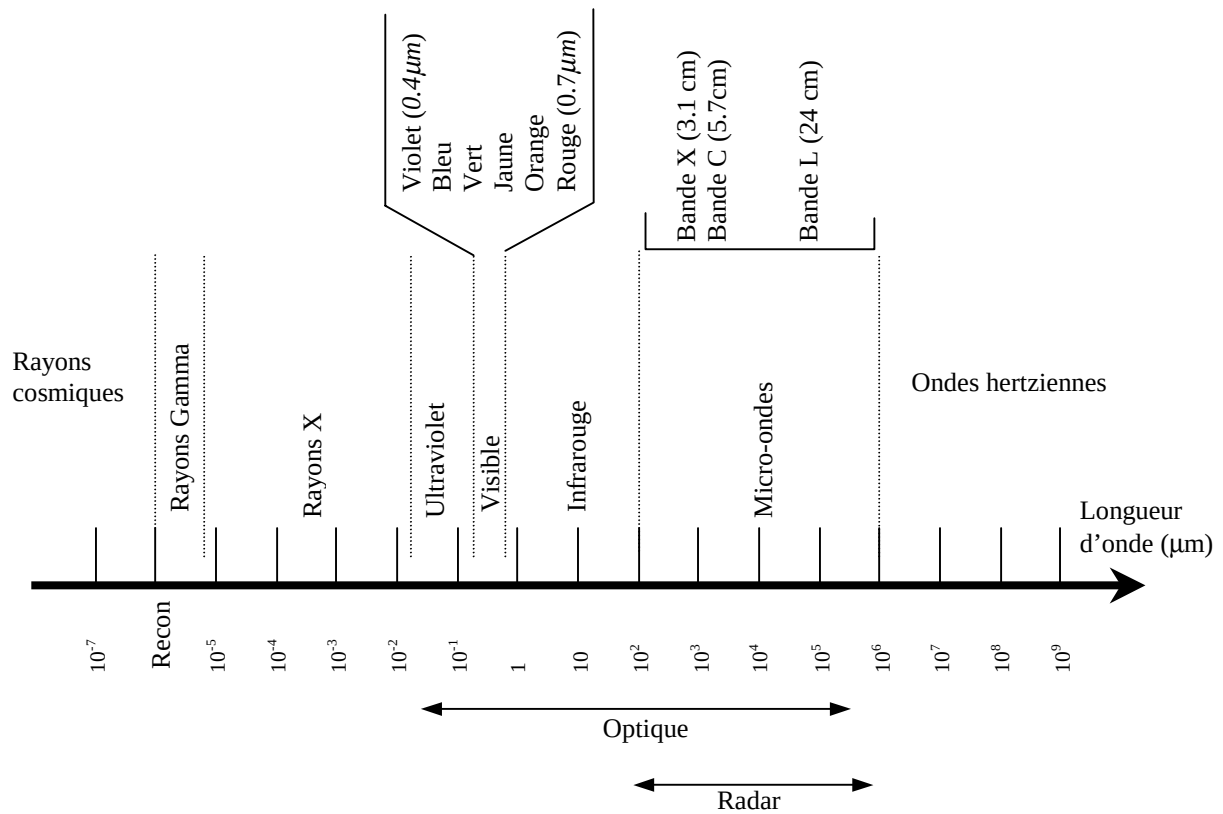


Figure 1. Spectre électromagnétique.

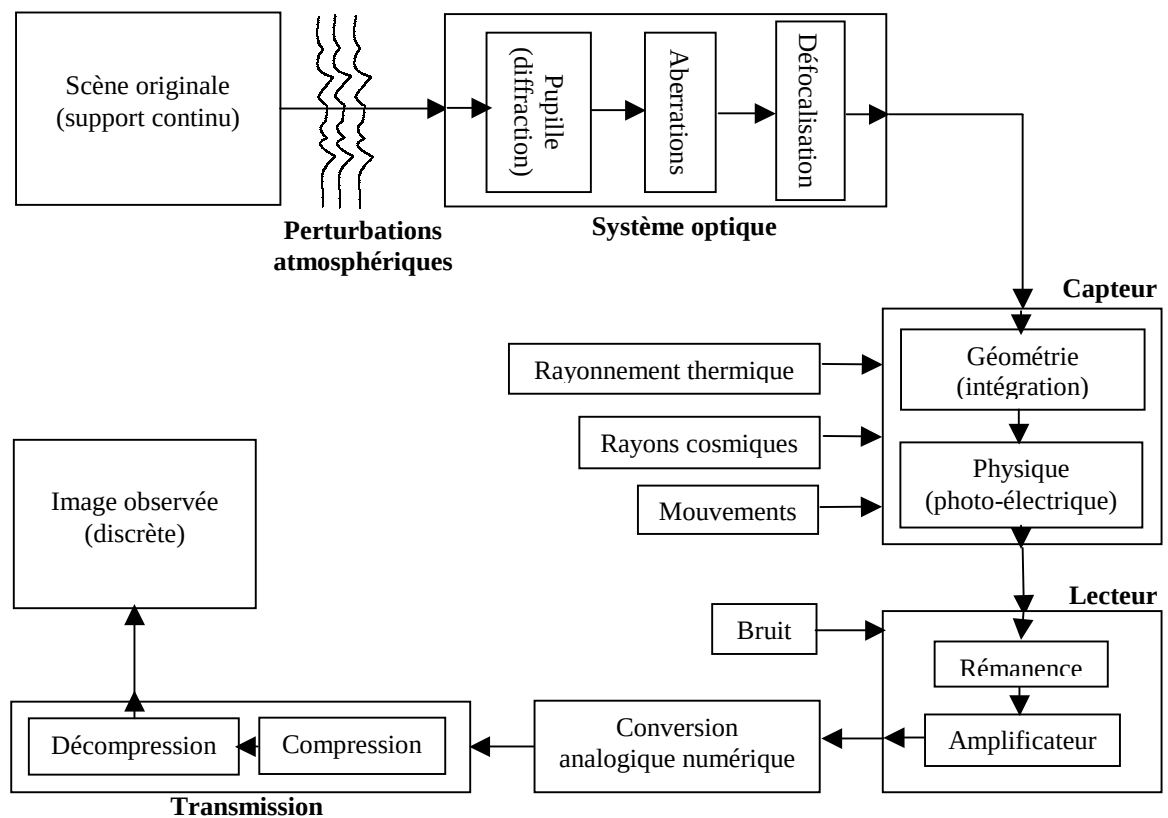


Figure 2. Chaîne de formation de l'image optique discrète à partir de la scène originale.

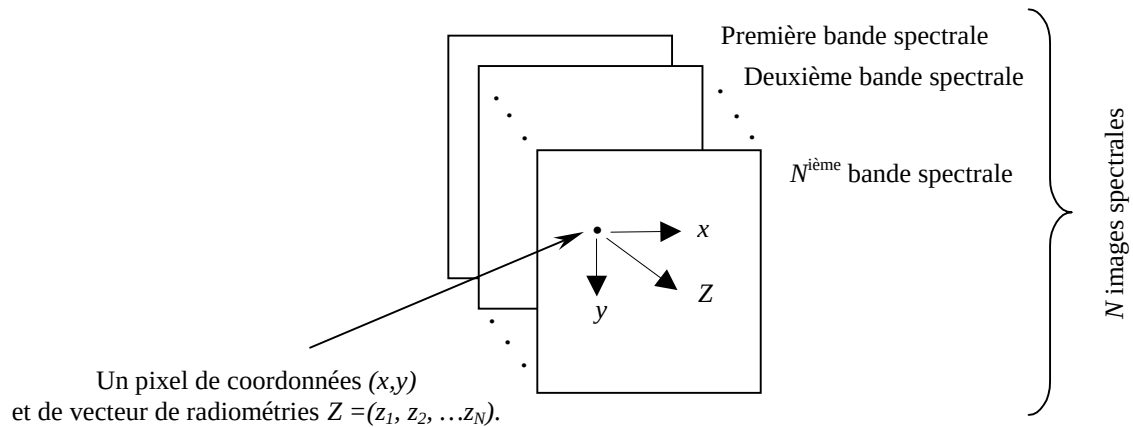


Figure 3. Image optique multispectrale fournie par le satellite.

Les lois de l'optique statistique nous permettent de décrire la dégradation de la scène observée lorsque les signaux réfléchis traversent d'abord l'atmosphère (où interviennent essentiellement les effets de la turbulence et de l'absorption), puis le système optique. Les signaux réfléchis subissent généralement une occultation liée à la taille finie de la pupille d'entrée et à la présence d'un miroir secondaire. C'est le phénomène de diffraction qui provoque la dégradation à ce stade.

Lors de la traversée de l'optique, l'on note aussi un certain nombre d'altérations qui peuvent être regroupées sous le nom d'aberrations, liées à certaines imperfections du système (aberrations systématiques ou erreurs diverses comme par exemple les défauts de polissage). Viennent s'ajouter aussi les aberrations liées à une mauvaise focalisation. Finalement, sur le capteur, se forme une image d'intensité dégradée. Cette image est dite floue : plus la taille de l'instrument diminue, plus la turbulence ou les aberrations sont importantes, et moins on observe de détails sur l'image, autrement dit moins l'image contient d'information. De plus, nous n'avons pas accès à cette image, mais à une version dégradée, cette fois-ci par le processus d'observation, ou d'acquisition. Tout d'abord, l'intensité est intégrée au niveau des pixels (éléments sensibles du détecteur). C'est le processus d'échantillonnage : passée cette étape, nous n'avons plus accès qu'à un ensemble discret de valeurs d'intensité intégrée. Cette étape contribue également au flou, non seulement à cause de l'extension spatiale des pixels, mais également par la durée de l'intégration, pendant laquelle le capteur se déplace et subit d'éventuelles vibrations.

Ensuite, cette intensité donne lieu à un déplacement d'électrons (effet photoélectrique). Le nombre d'électrons venant s'accumuler dans un puits de potentiel est une fonction idéalement linéaire de l'intensité intégrée. La sensibilité varie légèrement d'un pixel à l'autre, ce qui constitue également une forme de dégradation. Sans oublier les pertes importantes intervenant à ce stade, le rendement quantique des détecteurs dépasse rarement 40%, la majorité des photons ne sont jamais détectés.

Tableau 1. Classification des différentes dégradations subies par la scène originale dans le processus de formation d'image de télédétection optique.

Type De Dégradation	Flou	Bruit	Défauts Radiométriques
Atmosphère	• Turbulence • Absorption		Transmission
Optique	• Diffraction : pupille d'entrée • Aberrations : défauts systématiques et stochastiques • Défocalisation		
Capteur (géométrie)	• Intégration sur les pixels • Perte de sensibilité entre les pixels • Bougé pendant la durée de l'intégration		Sensibilité
Capteur (physique)	• Diffusion des charges • Lecture : effet de rémanence	• Photons d'origine thermique • Rayons cosmiques	• Non-stationnarité (sensibilité des pixels) • Non-linéarité (effet photoélectrique)
Electronique de conversion		• Bruit de mesure • Quantification	Non-linéarité (amplification)
Transmission bord-sol		• Compression/décompression • Pertes en transmission	

D'autre part, l'intensité intégrée utile n'est pas la seule à provoquer un déplacement d'électrons, viennent s'ajouter notamment des électrons d'origine thermique. D'autre part, l'isolement n'étant jamais parfait, ces électrons ont tendance à diffuser pour contaminer les pixels voisins. La phase de lecture du capteur consiste à récupérer les électrons accumulés pour les convertir en courant mesurable ; intervient donc une phase de transfert d'un pixel à l'autre qui peut avoir pour effet de mélanger un peu les charges accumulées, par un effet de rémanence. Le courant produit est amplifié, et donc sujet à une nouvelle non-linéarité ainsi qu'à une contamination par du bruit d'origine électronique. Finalement, le signal produit est numérisé, c.-à-d. quantifié sur un certain nombre de bits (de 8 à 16 en général), ce qui introduit une incertitude supplémentaire. Enfin, une nouvelle dégradation est appliquée par une chaîne de compression/ décompression. On obtient alors l'image finale, discrète, de taille finie, floue et bruitée. C'est sur cette image que nous allons travailler pour tenter d'obtenir une description aussi fidèle que possible de la scène originale.

Cependant, malgré la complexité de la chaîne de formation de l'image optique, il est possible de définir une chaîne équivalente plus simple qui nous permettent de classer les principaux types de dégradation non par ordre d'intervention, mais selon leur mode de

modélisation [79], [80]. En effet, nous distinguons quatre grands groupes d'opérateurs permettant de résumer le passage de la scène à l'image discrète observée. La chaîne simplifiée, illustrée par la figure 4, peut se résumer ainsi : application d'un flou, discrétisation spatiale, transformation des valeurs obtenues, puis addition de bruit. Ainsi, le processus de formation d'un pixel de coordonnées (x,y) de l'image optique, faisant intervenir les dégradations les plus courants et supposant que le bruit est aléatoire de caractère additif et survient indépendamment du signal image, peut être modélisé par un système linéaire invariant par translation, représentée par la figure 5, ayant la forme :

$$g(x,y) = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} f(m,n)h(x-m,y-n) + n(x,y) \quad (1)$$

où $f(m,n)$ est l'image idéale, $h(x-m,y-n)$ est le terme de dégradation, $n(x,y)$ est le bruit et $g(x,y)$ est l'image finale dégradée.

2. 2. Formations des Images Radars

L'image radar est le résultat de l'observation de la terre par un capteur actif embarqué sur un engin mobile (avion ou satellite). Pour cela, différents traitements complexes sont nécessaires pour transformer les données radars brutes en données images [81], [82], [6].

Le principe de formation des données radars en données images est présenté par la figure 6 [83]. Une antenne embarquée à bord de l'engin émet un train d'ondes électromagnétiques dans la bande micro-onde puis collecte les échos renvoyés par la portion de la surface terrestre. Généralement, la même antenne est utilisée en émission et en réception grâce à un dispositif (rotateur) qui permet à l'antenne de commuter d'un mode à l'autre. Il faut signaler ici que le rayonnement produit par l'antenne se fait de manière inclinée et non perpendiculaire au sol pour éviter que deux points symétriques par rapport à la perpendiculaire ne soient confondus en un seul point dans le système imageur.

Deux informations sont véhiculées par l'écho radar (amplitude et phase), la portée de l'objet, appelé aussi cible, et sa réflectivité radar [84]. La portée de la cible est la distance séparant capteur-cible. Celle-ci peut être calculée en mesurant le temps que met le signal radar pour faire le trajet aller-retour. La distance sera alors proportionnelle à ce temps et à la vitesse de la lumière. Par contre, la réflectivité radar mesure l'énergie réfléchiée par la cible. Comme tous les objets ne réfléchissent pas les ondes radar avec la même efficacité, on associe à toute cible une section efficace de rétrodiffusion radar. Celle-ci est d'autant plus élevée que le pouvoir réflecteur de cette cible est grand. On la définit comme le rapport entre l'énergie de l'écho et celle que recevait le radar si la cible était un diffuseur parfaitement isotrope.

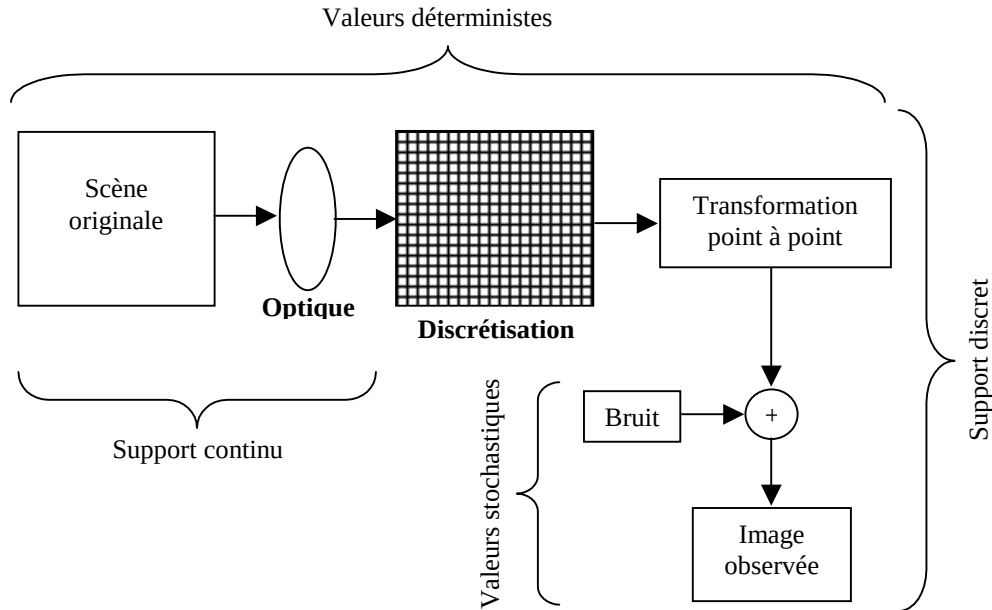


Figure 4. Chaîne simplifiée de formation de l'image optique discrète à partir de la scène originale.

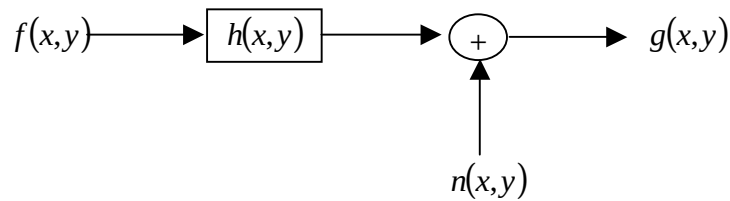


Figure 5. Modèle simplifié de formation de l'image optique.

La transformation des données radars en image exige un traitement mathématique lourd et complexe appelé focalisation, compression et reconstitution liés à la synthèse d'ouverture [81], [82], [6]. C'est seulement après ces traitements que les images sont prêtes à subir les traitements classiques (corrections géométriques et radiométrique) et interprétation thématique rencontrés en imagerie optique. Les radars dotés de cette capacité sont appelés radars à ouverture synthétique "*Synthetic Aperture Radar (SAR)*".

La mesure radar complète est la combinaison des polarisations linéaires horizontale (H) et verticale (V) à l'émission de l'onde radar et à la réception de l'onde rétrodiffusée. La mesure permet donc de restituer après étalonnage une matrice de rétrodiffusion dont les éléments sont des valeurs complexes et permettent de mettre en relation l'onde rétrodiffusée à l'onde incidente [83], [84]. Les radars conventionnels ne mesurent qu'une polarisation (en amplitude), c.-à-d. uniquement le module d'un seul des coefficients de la matrice de rétrodiffusion. Ils utilisent la même configuration de polarisation d'antenne à la fois pour l'émission et la réception du signal, et opèrent à une polarisation fixée.

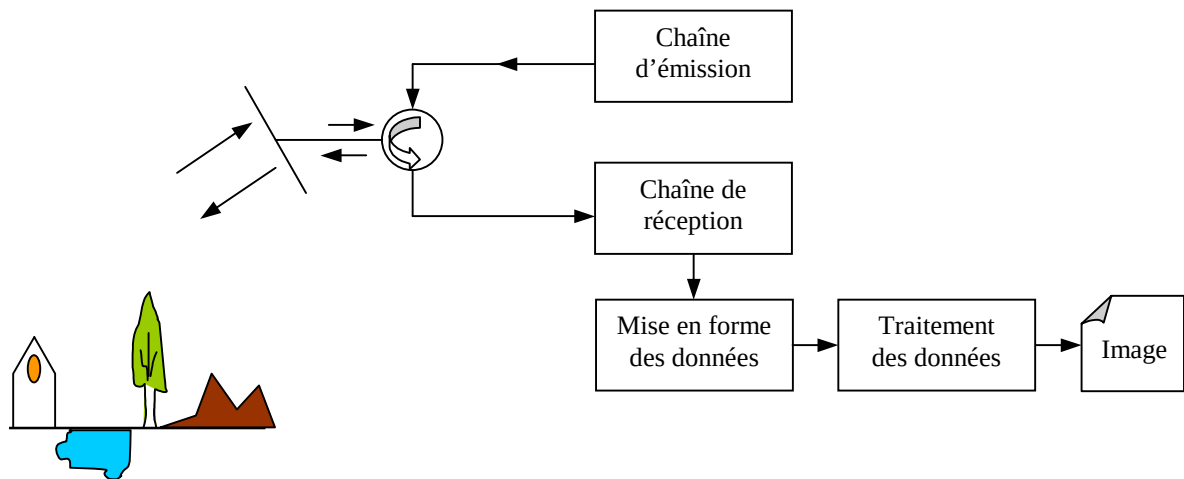


Figure 6. Principe de formation de l'image radar.

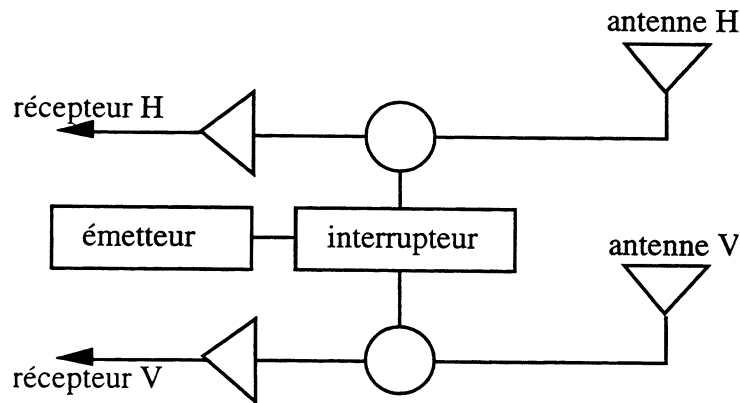


Figure 7. Configuration du radar AirSAR.

Le radar conventionnel ne permet donc pas une caractérisation complète des mécanismes d'interaction de l'onde électromagnétique avec la surface et les propriétés de dépolarisation des cibles naturelles ne peuvent être étudiées. On a donc conçu récemment des radars qui acquièrent toute l'information contenue dans le phénomène de rétrodiffusion, en mesurant, pour élément de la scène, l'ensemble des quatre polarisations de la réponse, en amplitude et en phase relative, c.-à-d. les matrices de rétrodiffusion complexes complètes. Ainsi, le radar AirSAR (du “*Jet Propulsion Laboratory (JPL)*” de la NASA “*National Aeronautics and Space Administration*”), par exemple, possède deux antennes émetrices pour chacune des polarisations H et V , ainsi que deux récepteurs H et V , dont la figure 7 présente un schéma simplifié. Le radar émet alternativement des impulsions en polarisation H ou V , tandis que l'onde rétrodiffusée est reçue simultanément en polarisation H et V . L'ouverture synthétique est réalisée par intégration cohérente, pour chacune des polarisations HH , HV , VH et VV , sur environ 1500 impulsions dans le cas des données AirSAR monovues ayant une résolution azimutale environ égale à mètre.

La section efficace radar est souvent utilisée pour mesurer les caractéristiques d'une cible simple. Elle s'exprime en fonction des éléments de la matrice de rétrodiffusion. Dans le cas d'une surface aléatoire, le coefficient de rétrodiffusion défini comme étant la section efficace radar moyenne par unité de surface est utilisé. Il fournit de ce fait une information sur les propriétés géométriques et diélectriques dépendantes de la fréquence d'émission, de son angle d'incidence, de la configuration de polarisation et de l'humidité de la cible [83].

Cette technique d'imagerie permet de produire des images de la terre de très haute résolution spatiale par un traitement des signaux reçus. La figure 8 illustre des résolutions en site et en azimuth d'un SAR. Le faisceau de l'antenne radar d'altitude h , de forme conique, d'axe formant un angle d'incidence θ avec la verticale (angle de visée), délimite sur le sol, à tout instant, une zone ellipsoïde inscrite dans un rectangle de cotés ΔX et ΔY .

Une fine résolution en site peut être obtenue en émettant une suite d'impulsions de même durée. Deux échos provenant de cibles différentes pourront alors être distingués s'ils sont séparés d'au moins d'une durée d'impulsion (si les cibles sont séparées en distance oblique dr). La résolution au sol dans la direction du site est dx . Les radars à ouverture synthétique se différencient des radars classiques, dits à ouverture réelle, par la possibilité d'obtenir également une forte résolution dans la direction azimuthale dy . La synthèse d'ouverture utilise le fait que le déplacement du radar induit un effet Doppler à partir duquel la résolution azimuthale est déduite [86]. La haute résolution en distance (perpendiculaire à l'axe du vol) est inversement proportionnelle à la largeur de bande émise alors que celle en azimuth (parallèle à l'axe du vol) est pratiquement la moitié de la longueur réelle en azimuth de l'antenne [82].

Une conséquence de l'utilisation de ce type d'imagerie est la sensibilité du capteur à l'interférence des ondes rétrodiffusées inhérentes à la cohérence du signal radar. D'une manière plus précise, la formation de chaque pixel de l'image SAR est le résultat de l'interférence (somme cohérente) des échos renvoyés par les diffuseurs constituant la cible, interférence qui peut être constructive (addition) ou destructive (soustraction) selon les valeurs des phases relatives de ces échos. Il s'ensuit que même une scène physiquement homogène apparaît, dans l'image, comme non homogène. Les pixels pour lesquels l'interférence est destructive apparaissent sombres tandis que ceux pour lesquels l'interférence est constructive apparaîtront clairs. Cette figure d'interférence confère une apparence bruitée à l'image : c'est le phénomène de speckle [85].

Pour modéliser le speckle, on considère que l'imageur radar est un système linéaire, invariant spatialement, qui peut être caractérisé en tout point par une réponse impulsionnelle $h(x,y)$ définissant les cellules de résolution. Le schéma simplifié de la transformation de l'onde rétrodiffusée en un pixel de l'image radar est représenté sur la figure 9 [90].

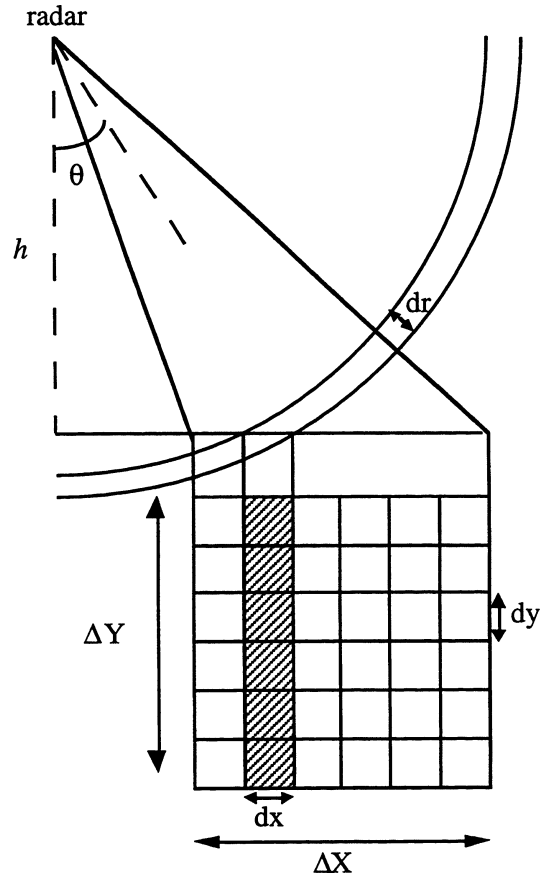


Figure 8. Illustration des résolutions en site et en azimuth d'un SAR.

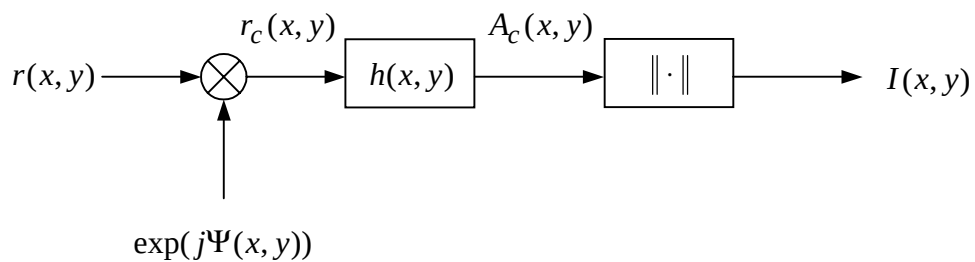


Figure 9. Modèle simplifié de formation d'image radar.

- $r(x, y)$: Amplitude ou réflectivité radar réelle de l'onde cohérente rétrodiffusée par la cible de coordonnées (x, y) et proportionnelle au coefficient de rétro-diffusion.
- $\psi(x, y)$: Phase aléatoire due aux différents diffuseurs constituant la cible et au trajet aller-retour de l'onde cohérente émise par le système radar.
- $I(x, y)$: Image formée en intensité.
- $h(x, y)$: Réponse impulsionnelle de l'imageur radar.

L'intensité produit pour chaque pixel de l'image radar peut être décrite par l'équation :

$$I(x,y) = \|A(x,y)\|^2 = \|h(x,y) * r_c(x,y)\|^2 \quad (2)$$

Comme :

$$r_c(x,y) = \sqrt{r(x,y)} \cdot \exp(j \cdot \psi(x,y)) \quad (3)$$

Il en résulte :

$$I(x,y) = \|h(x,y) * \sqrt{r(x,y)} \cdot \exp(j \cdot \psi(x,y))\|^2 \quad (4)$$

Lorsqu'une surface est physiquement homogène, la réflectivité radar est considérée comme constante. Il est donc possible de l'extraire du produit de convolution.

$$I(x,y) = r(x,y) \|h(x,y) * \exp(j \cdot \psi(x,y))\|^2 \quad (5)$$

En posant :

$$n(x,y) = \|h(x,y) * \exp(j \cdot \psi(x,y))\|^2 \quad (6)$$

Nous aboutissons à l'expression suivante :

$$I(x,y) = r(x,y) n(x,y) \quad (7)$$

Cette équation traduit le caractère multiplicatif du speckle $n(x,y)$ valable uniquement dans les zones homogènes ou faiblement hétérogènes. Les variations de l'intensité dans ces zones sont dues au seul speckle. Les propriétés statistiques associées à l'intensité seront celles du speckle dit totalement développé [91].

Le speckle peut être atténué avant ou après formation de l'image. L'atténuation du speckle avant formation de l'image s'effectue dans le domaine temporel ou fréquentiel [6] reposant sur le principe suivant : on atténue le speckle en additionnant deux ou plusieurs vues (looks en anglais) cohérentes indépendantes de la même scène. Cette technique, appelée traitement multi-vues, se fait en général à bord de l'engin mobile. Généralement un traitement de 3 ou 4 vues est largement suffisant pour atténuer son effet. Toutefois cette technique engendre une dégradation de la résolution spatiale d'un facteur N représentant de vues.

L'atténuation du speckle après formation de l'image consiste à atténuer le speckle dans le domaine spatial. Des méthodes plus robustes sont alors développées afin d'améliorer la qualité de l'image radar sans réduction de la résolution spatiale. Un état de l'art sur les différentes méthodes de filtrage du speckle dans le domaine spatial est donné dans [26].

2. 3. Influence de l'Atmosphère

L'interaction des rayonnements réfléchis par l'objet terrestre avec les constituants de l'atmosphère se traduit par une absorption ou une diffusion dans des bandes spécifiques. Pour illustrer ce phénomène, considérons le cas de l'atmosphère dont la transmission en fonction de la longueur d'onde est représentée sur la figure 10.

Les zones grises indiquent l'opacité de l'atmosphère (absorption) vis à vis du rayonnement; tandis que les zones blanches sont les fenêtres atmosphériques pour lesquelles l'interaction avec les constituants de l'atmosphère est quasi négligeable. Dans l'Ultraviolet, l'opacité de l'atmosphère est due essentiellement à la couche d'ozone. Par contre, dans le Visible et le Proche Infrarouge, il existe des bandes dans lesquelles l'atmosphère possède une opacité élevée due principalement à la vapeur d'eau et au dioxyde de carbone. Dans la bande Infrarouge lointain, l'atmosphère est totalement opaque; tandis que dans la bande micro-onde, l'atmosphère est transparente.

La présence des nuages conduit à une opacité additionnelle, associée à l'absorption et à la diffusion du rayonnement par les gouttelettes d'eau présentes dans les nuages. Cet effet se manifeste dans les domaines Visible et Infrarouge. Par contre, les nuages sont pratiquement transparents dans le domaine des micro-ondes ce qui constitue l'un des avantages de l'observation par un capteur radar.

Il faut signaler ici que le soleil constitue la seule source d'éclairement des objets terrestres dans les domaines du Visible et de l'Infrarouge. Par contre, dans le domaine des micro-ondes, la source d'éclairement est une onde radar générée à partir d'un système actif embarqué sur le satellite d'observation de la terre.

Les systèmes d'acquisition embarqués dans les satellites de télédétection civile sont conçus en faisant un choix judicieux des parties du spectre électromagnétique dans lesquelles l'opacité de l'atmosphère est négligeable. Ainsi, les principales bandes spectrales dans lesquelles des mesures sont possibles sont les suivantes :

- Le Visible, de 0.4 à $0.7\mu m$, souvent décomposé en bandes spectrales plus étroites.
- Le Proche Infrarouge, de 0.7 à $1.5\mu m$.
- Le Moyen Infrarouge, de 1.5 à $3\mu m$.
- L'Infrarouge Thermique, de 8 à $15\mu m$.
- Les micro-ondes, de $1mm$ à $1m$.

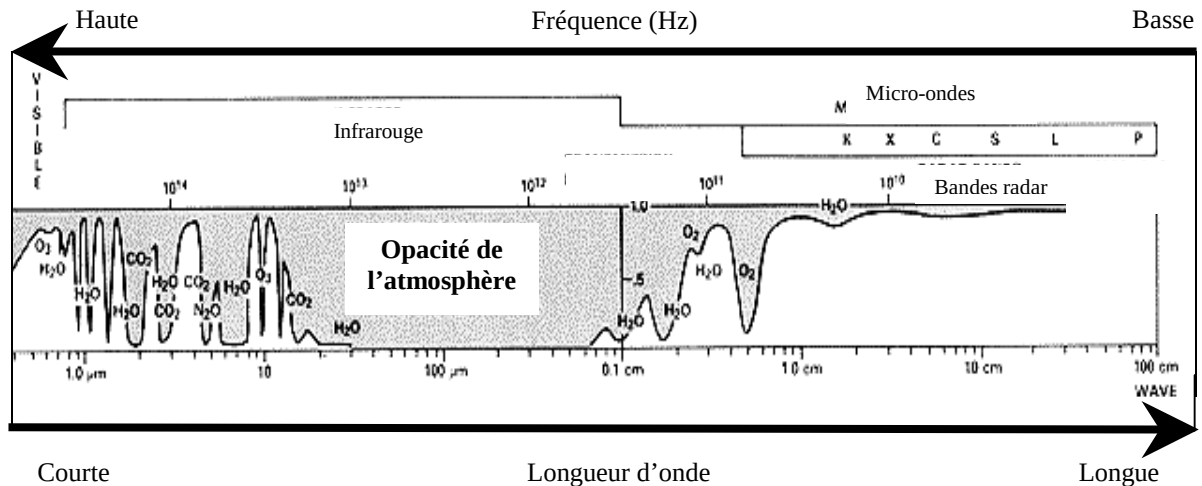


Figure 10. Spectre de transmission de l'atmosphère terrestre.

2. 4. Distorsions des Images de Télédétection

Les capteurs embarqués sur les satellites de télédétection civile fournissent des images de la terre dont les pixels subissent diverses distorsions notamment la modification de leur niveau de gris (distorsions radiométriques) et leur mauvaise position (distorsions géométriques). Des corrections doivent être alors appliquées à ces images afin d'éviter toute perte de précision pouvant réduire leur utilité lors de l'extraction des informations.

2. 4. 1. Distorsions Radiométriques

La modification de la radiométrie d'un pixel est dépendante, d'une part, des caractéristiques du capteur imageur (optique ou radar) et de son mode d'acquisition et, d'autre part, des conditions de prise de vue dans lesquelles l'image a été acquise.

Les distorsions radiométriques peuvent être séparées en deux catégories : celles qui affectent l'image optique et celles qui affectent l'image radar. Les radiométries dans l'image optique sont essentiellement modifiées par le passage de la radiation réfléchie à travers l'atmosphère. Cependant, une mauvaise calibration des détecteurs du capteur et les effets de l'éclairement du soleil constituent également des sources de distorsion [25]. Ces distorsions radiométriques sont considérées du point de vue de traitement d'images comme étant un bruit additif qui se rajoute à la valeur radiométrique du pixel.

Dans l'image radar, la principale distorsion radiométrique, appelée speckle, provient de l'interférence des ondes réfléchies liée au caractère cohérent du signal radar. En effet, lorsque l'onde radar est émise en direction de la cible, celle-ci renvoie non pas une seule

onde réfléchi mais plusieurs ondes qui interfèrent au niveau du capteur. Le speckle est un phénomène qui se produit chaque fois qu'il y a des interférences entre les réponses réfléchies. Il a les caractéristiques d'un bruit multiplicatif qui nuit à la lisibilité de l'image et rendre son exploitation souvent difficile.

2. 4. 2. Distorsions Géométriques

Les distorsions géométriques sont dues, d'une part, aux conditions de prise de vue et, d'autre part, aux résolutions spatiales différentes entre les capteurs. Elles peuvent être classées en deux catégories : les distorsions systématiques ou prédictibles et les distorsions essentiellement aléatoires [26]. Les distorsions systématiques sont parfaitement connues et peuvent être corrigées par l'application d'un ensemble de formules déduites au travers d'une modélisation mathématique. Tandis que les distorsions aléatoires sont corrigées grâce à l'élaboration d'un modèle d'approximation polynomiale utilisant des points de contrôle repérés sur l'image et sur une référence terrain ou cartographique. Les sources de distorsions géométriques communes aux images optiques et radars sont principalement :

- Rotation de la terre durant l'acquisition de l'image.
- Courbure de la terre.
- Effets liés à la scène et au capteur (effets panoramiques et de perspective).
- Mouvements de la plateforme du satellite (altitude, attitude et vitesse).
- Effets liés au relief terrestre.

Les effets liés au relief terrestre provoquent des erreurs géométriques spécifiques inhérentes au système d'imagerie radar pouvant affecter considérablement la qualité de l'image. La figure 8 représente, dans le plan d'élévation, un relief constitué de deux segments, de mêmes longueurs (segment 1-2 et segment 2-3) mais de pentes différentes. Si le radar est situé vers la gauche, le segment 1-2 sera imagé sur un plus petit nombre d'échantillons que le segment 2-3 : la colline apparaîtra donc raccourcie du côté du radar et étirée de l'autre.

Cet effet est appelé raccourcissement ou étirement (en anglais : *forshortening*) selon que la pente est située en face ou à l'opposé du radar; il se présente dans les zones où l'angle d'incidence local est proche de 90° . Il en résulte une compression de l'information radiométrique rétrodiffusée dans les zones situées en face du radar. Il peut être corrigé géométriquement si l'on dispose d'une carte topographique de la région. Par contre, lorsque la pente locale est supérieure à l'angle d'incidence du faisceau radar, le sommet de la colline (segment 3-4) sera imagé avant la base du fait que la distance radar-cible est plus courte. Ce phénomène se substitue à l'effet de raccourcissement et provoque une inversion ou un basculement (en anglais : *layover*) de l'information rétrodiffusée au niveau du capteur radar.

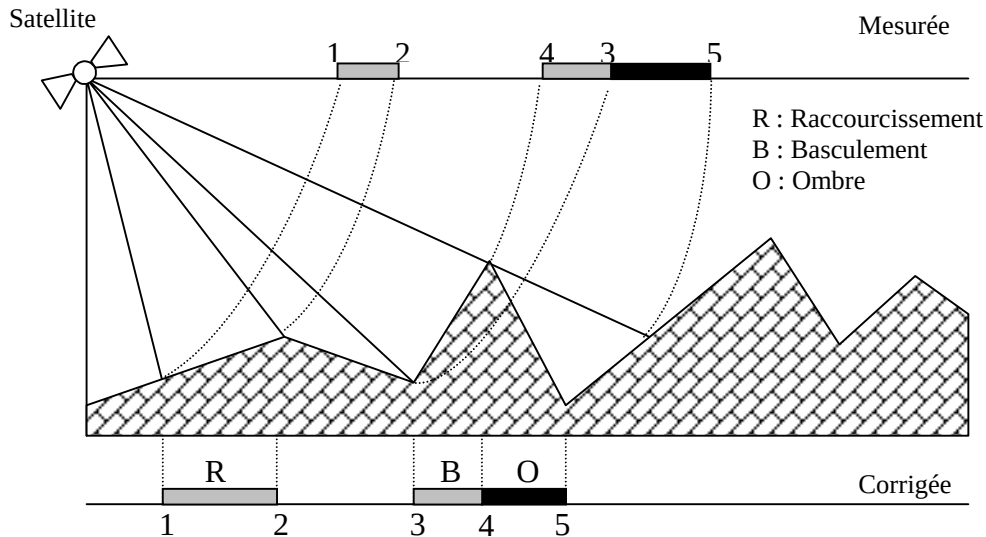


Figure 11. Erreurs géométriques introduites par les variations du relief.

Sur l'image, l'effet de basculement, impossible à corriger (sans l'utilisation d'une carte topographique), conduit à générer des pixels de hautes intensités apparaissant sous formes de points très brillants. D'autre part, le sommet d'une colline peut provoquer une absence de rétrodiffusion de la pente opposée au radar (segment 3-4). Ce phénomène d'ombre (en anglais : *shadowing*) se manifeste dans les zones non éclairées par le faisceau radar. Il en résulte une apparition de points noirs dans l'image.

2. 5. Contenu Informatif d'une Image de Télédétection

Le contenu informatif d'une image est donc conditionné par un certain nombre de variables qui se rattachent à quatre aspects : spectral, spatial, radiométrique et temporel.

2. 5. 1. Aspect Spectral

Le rayonnement émis, réfléchi ou diffusé par un objet fournit des informations sur les caractéristiques de la matière dont la nature dépend fortement de la longueur d'onde [11]. Chaque objet observé possède une « empreinte » caractérisée par une signature spectrale. Une signature spectrale est définie par une courbe de réflectance (pourcentage d'énergie réfléchie) de l'objet en fonction de la longueur d'onde. Ainsi, la mesure des propriétés de rayonnement de l'objet observé requiert l'association d'un capteur pouvant intercepter une partie du champ électromagnétique au travers des bandes spectrales. Le nombre et la nature de ces bandes définissent la résolution spectrale du capteur. Celle-ci mesure le pouvoir de discrimination du capteur en fonction de la longueur d'onde. Par conséquent, la sélection des bandes ou combinaison de bandes les plus appropriées requiert un soin particulier dans

la conception du capteur afin de pouvoir discriminer les signatures spectrales des matériaux constituant la surface terrestre.

2. 5. 2. Aspect Spatial

La richesse du contenu informatif d'une image dépend aussi de la capacité du capteur à discriminer deux objets voisins au travers de la caractérisation d'une résolution spatiale. Celle-ci est définie comme étant la plus petite distance en-deça de laquelle deux objets ne peuvent plus être spatialement séparés par le capteur et dépend donc de plusieurs facteurs (champ de visée instantanée du capteur, hauteur du satellite, taille des capteurs,...). Le pouvoir de résolution du capteur est donc d'autant meilleur que cette distance est petite.

2. 5. 3. Aspect Radiométrique

Le nombre de niveaux de gris d'une image, couvrant une plage de brillance donnée, définit la résolution radiométrique du capteur. Ce dernier doit posséder la capacité de discriminer deux objets de brillances voisines. Alors que l'œil humain est limité à environ 16 niveaux de gris, les images numériques en offrent typiquement 256 niveaux de gris (65536 niveaux de gris dans le cas de l'image radar).

2. 5. 4. Aspect Temporel

La signature spectrale d'un objet dépend non seulement de la longueur d'onde mais aussi du temps et de la localisation géographique. En effet, les propriétés spectrales des objets sont modifiées par les changements saisonniers qui doivent être pris en compte pour déterminer le moment le plus propice à l'acquisition des images dans le but de la réalisation d'une étude. Par exemple, le suivi de l'évolution de la couverture végétale d'un site géographique donné nécessite l'acquisition de plusieurs images prises à différentes périodes de l'année, ou à la même date mais à différentes années. Ces images sont ensuite analysées [12], [13] pour déterminer et quantifier le type et le lieu des modifications des réponses spectrales.

2. 6. Capteurs Imageurs en Télédétection Spatiale Civile

Actuellement, de nombreux capteurs de télédétection spatiale civile permettent de fournir des images de la terre de différentes résolutions spectrale et spatiale. Ces capteurs sont conçus en fonction des objectifs définis par les applications civiles. C'est pourquoi, nous pouvons distinguer deux classes de capteurs appelés aussi imageurs à savoir : imageur optique et imageur radar.

2. 6. 1. Imageurs Optiques

Dans l'imagerie optique, qui correspond à la partie optique du spectre électromagnétique, à savoir les gammes : Ultraviolet, Visible, Infrarouge, et Proche Infrarouge, les capteurs photoélectriques convertissent les radiations réfléchies par les objets terrestres en signaux électriques directement utilisables pour la transmission au sol en temps réel. Ces signaux sont traduits sous forme d'images à exploiter. Nous distinguons trois types de capteurs optiques : capteurs spectraux, capteur panchromatique, et capteur thermique.

Les capteurs spectraux mesurent les réponses liées à la signature spectrale électromagnétique de la surface terrestre observée suite à un rayonnement solaire. Cette signature dépend essentiellement de la nature et de l'état de l'objet terrestre au moment de l'observation (paramètres de couleur et de température de l'objet terrestre) ainsi que des conditions de propagation des ondes électromagnétiques réfléchies par celui-ci et son environnement. Les images ainsi obtenues par ces capteurs sont appelées images multispectrales optiques. Le capteur panchromatique fournit une image dans le Visible et se distingue par sa haute résolution spatiale. L'image fournie par ce capteur est appelée image panchromatique et est destinée habituellement pour des applications cartographiques. Dans la gamme Infrarouge Thermique, le capteur thermique mesure les rayonnements émis par la terre elle-même et l'image obtenue est appelée image thermique.

La nouvelle technologie des systèmes optiques d'acquisition a permis de fournir des images multispectrales correspondant à des centaines de capteurs opérant dans la partie optique du spectre électromagnétique [14]-[17]. Le système aéroporté "*The airborne Visible/Infrared Imaging Spectrometer (AVIRIS)*", par exemple, fournit une image multispectrale correspondant à 224 capteurs. Les bandes spectrales correspondantes sont espacées à approximativement 10 nanomètres commençant à une longueur d'onde de 440nm. Chaque pixel a une précision radiométrique de 12 bits et une résolution spatiale de 20m. Cette amélioration dans la résolution spectrale devrait fournir une richesse d'informations détaillées sur la scène imagée.

La contrainte essentielle de l'imagerie optique réside dans la dépendance des conditions météorologiques (les nuages, la pluie, le brouillard) lors de l'observation d'un site géographique donné. Ainsi, elle ne permet pas un suivi régulier d'éventuels changements terrestres particulièrement dans les zones tropicales et les hémisphères nord et sud où la présence de la couverture nuageuse est quasi-permanente.

Une alternative à l'imagerie optique consiste en l'utilisation d'une source d'éclairage pouvant traverser les nuages. Cette source est générée par un capteur, embarqué sur le satellite, appelé radar opérant dans le domaine des micro-ondes.

2. 6. 2. Imageurs Radars

L'imageur radar est un capteur actif pouvant détecter les ondes réfléchies qu'il a lui-même générées. D'autre part, un choix judicieux de la longueur d'onde permet de rendre l'atmosphère totalement transparente vis-à-vis du signal radar. Ces deux caractéristiques font que l'imageur radar permet de réaliser des observations de jour comme de nuit indépendamment des conditions météorologiques [9], [18].

Depuis la conception de systèmes radar à ouverture synthétique SAR permettant l'acquisition d'images de forte résolution spatiale, de l'ordre de 10 mètres, l'utilisation des capteurs hyperfréquences à la télédétection connaît un essor grandissant, dont témoigne le développement de capteurs spatiaux tels que les satellites "*European Remote Sensing (ERS) Satellites*", RadarSAT, EnviSAT. Dans cette partie du spectre électromagnétique, les capteurs transmettent des signaux de fréquences, polarisations, et angles d'incidence spécifiques et mesurent ensuite les réponses réfléchies par la surface terrestre [9]. Cependant, dans la gamme des micro-ondes, la rugosité, les propriétés électriques, et la forme géométrique de l'objet terrestre influent sur les rayonnements réfléchis [4], [6], [18]. D'une manière plus précise, la formation d'un pixel de l'image SAR est le résultat de l'interférence (somme cohérente) des échos renvoyés (signaux complexes) par les diffuseurs constituant la cible visée, interférence qui peut être constructive (addition) ou destructive (soustraction) selon les valeurs des phases relatives de ces échos. Les pixels pour lesquels l'interférence est destructive apparaissent sombres tandis que ceux pour lesquels l'interférence est constructive apparaissent clairs. Cette figure d'interférence confère une apparence bruitée à l'image : c'est le phénomène de speckle.

Comme les capteurs optiques, les capteurs radars peuvent opérer à différentes fréquences et polarisations. En outre, ces derniers étant des systèmes actifs, il est possible d'émettre des signaux de fréquences et polarisations différentes et de recevoir des ondes selon différentes configurations de polarisation. Ainsi, la dernière décennie a vu le développement d'une nouvelle génération de radars multifréquences et multipolarisations, tels que AirSAR et SIR-C/X "*Spaceborne Imaging Radar C/X*", tous deux du JPL de la NASA [19], [20]. Les images ainsi obtenues sont appelées image SAR polarimétriques. L'apport d'informations fourni par ces nouveaux instruments, par rapport au cas de radars monofréquences et monopolarisations tels que ERS s'est d'ores et déjà avéré très efficace pour une meilleure caractérisation des interactions de l'onde électromagnétique avec la surface imagée.

L'image fournie par l'imageur radar doit subir un traitement mathématique complexe pour la transformer en une image de qualité similaire à celle d'une image optique [24]. Ce traitement fournit une image radar dont les intensités sont dépendantes des caractéristiques du signal d'éclairement tels que la longueur d'onde, le type de polarisation (horizontale ou

verticale), l'angle d'incidence et la trajectoire du satellite. La rugosité, la forme, l'orientation et le contenu en humidité du sol sont les principales informations contenues dans cette image.

2. 6. 3. Comparaison Entre Imageur Optique et Imageur Radar

Par rapport aux capteurs optiques, les systèmes radar hyperfréquences présentent deux avantages majeurs : tout d'abord, le radar est un système actif qui s'affranchit des variations du rayonnement solaire et garde l'avantage par sa capacité d'observation de jour comme de nuit ; ensuite, dans la gamme des fréquences utilisées, de 1 à 10GHz, les ondes électromagnétiques sont quasiment insensibles aux composantes de l'atmosphère.

Les inconvénients majeurs des capteurs optiques est qu'ils ne peuvent pas fournir des informations sur la surface imagée pendant la nuit ou bien si le ciel est nuageux. D'autant plus que l'atmosphère par ces phénomènes d'absorption et de diffusion, perturbe la réponse réfléchie par la surface. Cependant, contrairement à l'imagerie optique, le traitement et l'interprétation des images radars sont souvent difficiles car elle est influencée par plusieurs paramètres. La sensibilité de l'imagerie radar à la géométrie des cibles permet de fournir des images qui renseignent sur la rugosité, la topographie et l'humidité du sol, mais ceci est fortement dépendant du type de polarisation et de la longueur d'onde du signal émis.

Cependant, nous pouvons considérer que les images fournies par les capteurs optiques et radars contiennent des informations complémentaires sur la scène observée. Ainsi, l'exploitation simultanée des deux types d'images peut être d'intérêt important pour une caractérisation fidèle de l'état de surface de la scène observée.

2. 6. 4. Capteurs Imageurs Utilisés

Depuis 1972, date de lancement de premier satellite de télédétection civile, plusieurs capteurs imageurs ont été conçus pour l'observation terrestre. Nous présentons ici seulement les capteurs imageurs de télédétection dont les images fournies correspondantes ont été largement utilisées dans la pratique.

2. 6. 4. 1. Série des satellites Landsat :

Les capteurs des satellites de la série Landsat sont les radiomètres imageurs les plus utilisés pour la cartographie, notamment pour les études géologiques [21] et la végétation [22]. Le premier satellite de Landsat a été lancé en 1972 ("*Landsat MSS: MultiSpectral Scanner*"), et fut le premier système ayant fourni des images sur le Globe sur une base régulière et prévisible. Actuellement, Landsat "*Thematic Mapper (TM) 5*", de six bandes spectrales (du Visible et de l'Infrarouge (proche, moyen)) de résolution 30 mètres au sol et une bande

thermique de résolution 120 mètres au sol, est opérationnel. Landsat TM 7, récemment lancé, est également opérationnel. Il a huit bandes, six bandes spectrales de résolution 30 mètres au sol, une bande thermique de résolution 120 mètres au sol, et une bande panchromatique de résolution 15 mètres au sol. La surface complète de la terre peut être couverte par le satellite Landsat tous les 16 jours. Les caractéristiques générales des capteurs équipant les satellites de la série Landsat sont données dans le Tableau 2.

2. 6. 4. 2. Série des satellites SPOT :

La série de satellites SPOT « Système Probatoire d'Observation Terrestre » est réalisée par le CNES « Centre National d'Etude Spatial, France » en collaboration avec la Suède et la Belgique. Les trois premiers satellites, SPOT 1, 2, et 3, disposent de deux instruments HRV « Haute Résolution dans le Visible » pouvant fournir une image dans la bande panchromatique (Visible) ou trois images dans les bandes spectrales du Vert, Rouge et Proche Infrarouge. Ils présentent les particularités suivantes :

- Avantage pour la haute résolution spatiale de ces images pouvant être exploitées dans des applications de cartographie thématique [23].
- Possibilité de visée oblique à l'aide d'un miroir orientable par télécommande (l'angle d'orientation est choisi par l'utilisateur).
- Avantage de la visée latérale, qui réside dans la possibilité d'obtenir des couples d'images stéréoscopiques, puisqu'une même scène peut être prise sous des angles différents lors de révolutions orbitales successives.

Le tableau 2 résume les principales caractéristiques des satellites SPOT. Les capteurs de résolution spatiale très élevée sont maintenant disponibles pour des applications civiles. Le satellite SPOT 4, lancé en 1998, utilise un nouvel instrument appelé la géométrie de haute résolution "*High Resolution Geometry (HRG)*", qui peut fournir des images de résolution de 5m, comparé à la résolution de 10m du mode à haute résolution des satellites SPOT 1, 2, et 3. Le SPOT 5 est le dernier satellite d'observation de la terre réalisé. Dans la nuit du 3 au 4 mai dernier, SPOT 5 a été placé sur orbite depuis le Centre Spatial Guyanais de Kourou. Les caractéristiques de ce satellite sont nettement améliorées comparativement à SPOT 4 :

- Trois instruments de prise de vue embarqués : HRG, HRS "*High Resolution Stereoscopia*" et VEGETATION 2.
- Une meilleure résolution en mode panchromatique : 5 m et 2,5 m (au lieu de 10 m).
- Une meilleure résolution en mode multispectral : 10 m (au lieu de 20 m) dans les 3 bandes Visibles et dans le Proche Infrarouge. La bande moyen infrarouge conserve cependant une résolution de 20 m pour des raisons liées aux caractéristiques des capteurs utilisés dans cette gamme de longueur d'onde.
- La fauchée est toujours de 60 km pour chacun des instruments.

- La capacité de prise de vue oblique est conservée.
- Un instrument dédié à l'acquisition de vue stéréoscopique est embarqué.

Les premières images ont été reçues à Toulouse le 7 mai à 11h27, heure locale. Elles représentent la ville d'Athènes (Grèce) dans une résolution de 2,5 m obtenue à partir de deux images à 5 m grâce à la technique du Supermode.

2. 6. 4. 3. Instruments Végétation :

L'instrument VEGETATION-1 est un capteur embarqué dans le satellite SPOT 4 opérationnel depuis 1998. Il a été étudié spécialement pour le suivi de la végétation, notamment dans le domaine des productions agricoles, de la foresterie et de la surveillance de l'environnement terrestre à l'échelle mondiale. Cet instrument comprend un système imageur dans quatre bandes spectrales: Bleu (0,43-0,47 microns), Rouge (0,61-0,68 microns), Proche Infrarouge (0,78-0,89 microns, et moyen Infrarouge (1,58-1,75 microns). Le Rouge et le Proche Infrarouge sont particulièrement bien adaptés pour décrire l'activité photosynthétique de la végétation, tandis que le moyen Infrarouge est un bon détecteur de l'humidité du sol et de la végétation. Quant au Bleu, il est destiné dans ce cas à effectuer des corrections atmosphériques. VEGETATION-1 utilise des barrettes de détecteurs semblables à celles de SPOT HRVIR ("Haute Résolution Visible et Infra Rouge"), et des optiques télécentriques assurant une résolution spatiale quasi-constante à travers le champ d'ouverture (2 200 km au sol). La précision du positionnement absolu des pixels est inférieure à 350 m. Le décalage dans les superpositions multitudes est inférieur à 500 m. La résolution des images produites est 1 km, dans la projection plate carrée. L'instrument VEGETATION-2, embarqué dans le satellite SPOT 5, est opérationnel depuis mai 2002. Il a les mêmes caractéristiques que l'instrument VEGETATION-1.

2. 6. 4. 4. Système AirSAR :

Le radar aéroporté AirSAR du NASA/JPL, embarqué à bord d'un avion DC-8, permet d'acquérir des données totalement polarimétriques. Dans le cas de données 4-look, qui sont celles que nous utiliserons par la suite, la résolution au sol est de 12.6m en azimuth sur 12.6m en distance oblique. AirSAR collecte les données simultanément dans trois bandes de fréquences : P, L et C. Le tableau 2 résume les principales caractéristiques du système radar AirSAR.

2. 6. 4. 5. Système SIR-C/X SAR :

Le radar "Spaceborne Imaging Radar C/X" SIR-C/X de la NASA/JPL est le premier radar spatial capable d'acquérir des données en trois bandes fréquentielles, 1248,75 MHz, 5298,75 MHz, et 9,6 GHz (correspondant aux bandes L, C, X, respectivement) avec différents angles d'incidence variant entre 20° et 60°. Porté sur la navette spatiale Endeavor, le système SIR-C/X SAR a acquis deux ensembles de données en Avril et Octobre 1994 [19], [20]. Les données ont été simultanément acquises en trois bandes

fréquentielles, avec différents angles d'incidence. A l'angle d'incidence 44° , les données acquises sont totalement polarimétriques (Horizontal-Horizontal (*HH*), Horizontal-Vertical (*HV*), Vertical-Horizontal (*VH*) et Vertical-Vertical (*VV*)). Mais aux autres angles d'incidence, seules des mesures en polarisations *HH* et *HV* ont été acquises. Les données sont des images multivues complexes. L'espacement entre les pixels est de $12.5m$ dans les deux directions en lignes et en colonnes, la résolution spatiale étant de $25m$ dans les directions en site et en azimut.

3. Traitement d'Images de Télédétection Spatiale Civile

En raison des faiblesses inhérentes aux différents systèmes d'acquisition, optiques et radars, et la nature des images obtenues, ainsi que la quantité d'information acquise relative à une scène imagée, qui est considérable dans le cas des images optiques multispectrales et radars polarimétriques, l'exploitation directe des images de télédétection est impossible. La complexité des informations contenues dans ces images rend leur analyse par un photo-interprète lourde et ardue. Des traitements sont alors nécessaires pour pouvoir exploiter ces images.

Dans cette thèse, une méthodologie globale de traitement d'images de télédétection spatiale civile, multisources optiques et radars, est proposée. La méthodologie peut être utilisée dans l'importe quel domaine d'application. Elle est constituée des étapes suivantes : modélisation, analyse, classification et fusion. Pour un photo-interprète, et selon l'objectif pratique à atteindre, le traitement requis peut être composé d'une seule ou plusieurs étapes. Nous décrivons brièvement dans ce qui suit les différentes étapes. Ces étapes seront détaillées ensuite, le long de cette thèse, dans des chapitres séparés. Les méthodes de filtrage, de détection de contours et segmentation sont aussi données mais ne seront pas considérées dans le reste de la thèse.

3. 1. Modélisation d'Images de Télédétection

La caractérisation statistique et la modélisation d'images de télédétection sont d'une importance capitale dans le développement des algorithmes d'analyse et de classification. Ceci consiste à décrire les comportements statistiques des images fournies par les satellites de télédétection civile, optiques et radars, par des modèles probabilistes appropriés. L'exactitude des estimateurs des paramètres des modèles est un problème crucial qui dépend de plusieurs facteurs en particulier de l'opération d'extraction des échantillons d'apprentissage. Les différentes méthodes d'estimation des paramètres statistiques des modèles ainsi que les manières d'obtenir des estimateurs précis sont alors l'essieu de la modélisation des images.

Tableau 2. Caractéristiques des capteurs équipant la série de satellites Landsat.
(V : Vert, R : Rouge, B : Bleu, P. IR : Proche Infrarouge, M. IR : Moyen Infrarouge,
T : Thermique, Pan : Panchromatique).

	Résolution spectrale (µm)	Résolution spatiale (m)	Champ d'observation (km)
Landsat MSS	Bande 4 (V) : 0.50 – 0.60	79x79	185x185
	Bande 5 (R) : 0.60 – 0.70	79x79	185x185
	Bande 6 (P. IR) : 0.70 – 0.80	237x237	185x185
	Bande 7 (M. IR) : 0.80 – 1.10	79x79	185x185
Landsat TM 5	Bande 1 (B) : 0.45 – 0.52	30x30	185x185
	Bande 2 (V) : 0.52 – 0.60	30x30	185x185
	Bande 3 (R) : 0.63 – 0.69	30x30	185x185
	Bande 4 (P. IR) : 0.76 – 0.90	30x30	185x185
	Bande 5 (M. IR) : 1.55 – 1.75	30x30	185x185
	Bande 6 (T) : 10.40 – 12.5	120x120	185x185
	Bande 7 (M. IR) : 2.08 – 2.35	30x30	185x185
Landsat TM 7	Bande 1 (B) : 0.45 – 0.52	30x30	185x185
	Bande 2 (V) : 0.52 – 0.60	30x30	185x185
	Bande 3 (R) : 0.63 – 0.69	30x30	185x185
	Bande 4 (P. IR) : 0.76 – 0.90	30x30	185x185
	Bande 5 (M. IR) : 1.55 – 1.75	30x30	185x185
	Bande 6 (T) : 10.40 – 12.5	120x120	185x185
	Bande 7 (M. IR) : 2.08 – 2.35	30x30	185x185
	Bande 8 (Pan) : 0.50 – 0.75	15x15	185x185

Tableau 3. Caractéristiques des capteurs équipant la série de satellites SPOT
(SPOT 1, 2 et 3).

	Résolution spectrale (µm)	Résolution spatiale (m)	Champ d'observation (km)
Mode panchromatique	0.51 – 0.76	10x10	60x60
Mode multispectrale	XS1 : 0.50 - 0.59	20x20	60x60
	XS2 : 0.61 - 0.68	20x20	60x60
	XS3 : 0.79 - 0.89	20x20	60x60

Tableau 4. Caractéristiques du système radar AirSAR.

Caractéristiques	Bande P	Bande L	Bande C
Longueur d'onde (cm)	68.38	24.02	5.662
Fréquence central	9,6 GHZ	1248.75 MHZ	5298.75 MHZ
Durée de l'impulsion (µs)	10		
Période des impulsions (ms)	13		
Dimension des antennes (cm)	91.4x182.9	45.7x161.3	16.5x135.9
Nombre de looks	4		
Dimension d'un pixel en azimut (m)	12.20		
Dimension d'un pixel en site (m)	6.662		
Altitude nominale (m)	8052		
Angle d'incidence (Deg.)	24.0-62.6		

Pour les images optiques, le modèle Gaussien et la problématique de l'estimation de ces paramètres statistiques sont abordées. L'exactitude de l'estimation est fortement dépendante d'une part de la représentativité des échantillons d'apprentissage et d'autre part du nombre de ces échantillons. En pratique, l'extraction d'un nombre suffisant d'échantillons fiables est une tâche très fastidieuse. Les critères de sélection des échantillons d'apprentissage sont donnés ainsi que l'estimation des statistiques d'ordre un et deux. Le problème d'exactitude de l'estimation est soulevé.

Le cas des images multispectrales de dimension élevée a été traité en mettant l'accent sur l'importance de l'exactitude de l'estimateur de la matrice de covariance de la classe (représentant un objet terrestre). Ainsi, des méthodes proposées pour l'estimation de la matrice de covariance ont été étudiées tout en faisant ressortir les inconvénients et les avantages de chacune d'elles.

Pour la modélisation des images radars, plusieurs modèles probabilistes ont été proposés dans la littérature. Au début de l'exploitation des images SAR, les distributions telles que Lognormale, Weibull, Beta modifiée, et Gamma ont été testées. Des chercheurs ont néanmoins découvert que ces distributions ne décrivent pas convenablement les comportements des signaux radars réfléchis. Récemment, la distribution K et la distribution complexe de Wishart ont été proposées. Ici, dans cette thèse, nous avons insisté sur ces deux modèles. Nous avons montré que l'utilisation de la distribution de Wishart est restrictive dans la pratique pour des raisons qui ont été justifiées. Le développement théorique de la distribution K comme modèle pour les images radars ainsi que l'estimation de son paramètre ont été développés. D'autres modèles qui ont été proposés dans la littérature ne seront pas considérés en détail, mais ils ont été comparés avec la distribution K.

3. 2. Analyse Multidimensionnelle d'Images Optiques

Les images multispectrales fournies par les satellites peuvent subir des déformations et des perturbations dues aux effets atmosphériques, aux vibrations des capteurs du système d'acquisition, et autres. Elles sont alors affectées par un bruit additif.

Le rapprochement des parties du spectre électromagnétique correspondant aux capteurs du système d'acquisition se traduit par des images spectrales corrélées, constituant l'image multispectrale. Par exemple, nous présentons dans la figure 12 les courbes de réflectance, exprimées en pourcentage, en fonction de la longueur d'onde, exprimée en nanomètres (nm), pour mettre en évidence le recouvrement spectral des images spectrales et panchromatique du satellite SPOT.

Nous constatons que l'image panchromatique recouvre les deux premières bandes spectrales (Vert et Rouge) et une portion de la troisième bande (Infrarouge). Ce recouvrement se traduit par de fortes valeurs de corrélation entre l'image panchromatique et les images spectrales. Les deux bandes spectrales *XS1* et *XS2* présentent une zone de chevauchement variant de 600 à 650nm et la zone séparatrice entre les deux bandes spectrales *XS2* et *XS3* est presque inexistante. Ces rapprochements spectraux sont caractérisés par des valeurs de corrélation entre les bandes plus ou moins importantes. Ainsi, le traitement de toutes les images spectrales implique donc un certain degré de redondance. Si la taille de la scène observée et/ou le nombre d'images spectrales croît, l'espace mémoire pour le stockage des données devient important et le coût de traitement augmente. Donc, une analyse multidimensionnelle est nécessaire. Elle nous permet d'avoir une nouvelle représentation optimale des images de télédétection dans laquelle la corrélation entre les images, le nombre d'images et l'effet du bruit sont réduits (voir figure 13).

Les méthodes d'analyse multidimensionnelle d'images optiques proposées dans la littérature peuvent être divisées en deux catégories : Méthodes de sélection des caractéristiques discriminantes, en se basant sur le pouvoir discriminant entre les classes fourni par les images spectrales, et les méthodes d'extraction de caractéristiques via une transformation linéaire des images spectrales. Cette transformation linéaire peut être déduite par plusieurs approches [5], [7], [10], [15].

Dans cette thèse, seules les méthodes d'extraction de caractéristiques sont considérées. Dans ce contexte, les approches statistiques d'Analyse en Composantes Principales (ACP) et ses variantes appliquées aux images optiques de télédétection ont été développées. Les autres méthodes d'extraction de caractéristiques ne seront pas considérées mais le lecteur peut se référer aux références correspondantes citées dans la bibliographie du chapitre 3.

Les images multispectrales optiques de dimension élevée de la nouvelle génération des systèmes d'acquisition posent le problème des techniques et méthodes utilisées pour les traiter. Les méthodes et les techniques développées pour le traitement des images de dimension faible ne sont pas valables pour traiter les images de dimension élevée. Un autre problème posé est le temps de calcul qui augmente avec le nombre d'images spectrales traitées.

Ainsi, afin d'exploiter la richesse d'informations contenue dans ce type d'images, de nouvelles méthodes et algorithmes doivent être développés. Ainsi, de nouvelles variantes de l'ACP pour l'analyse d'images multispectrales de dimension élevée ont été étudiées dans cette thèse.

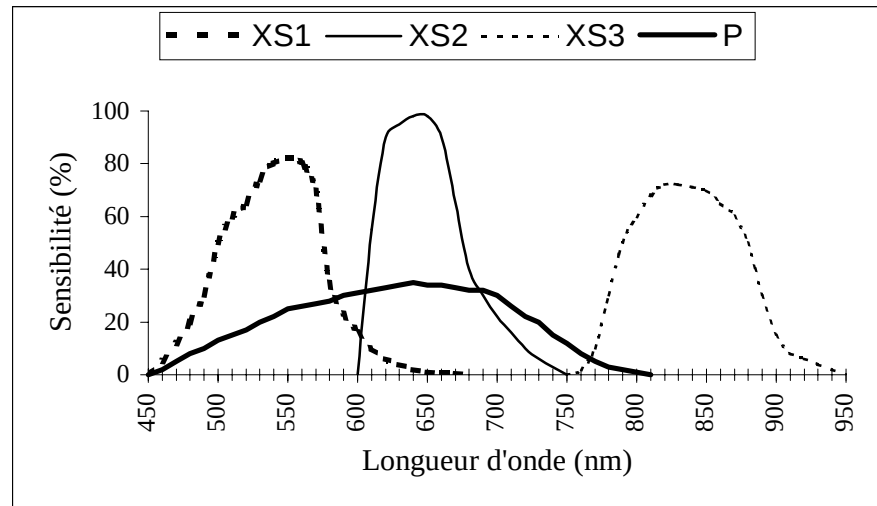


Figure 12. Courbes de sensibilité des capteurs SPOT.

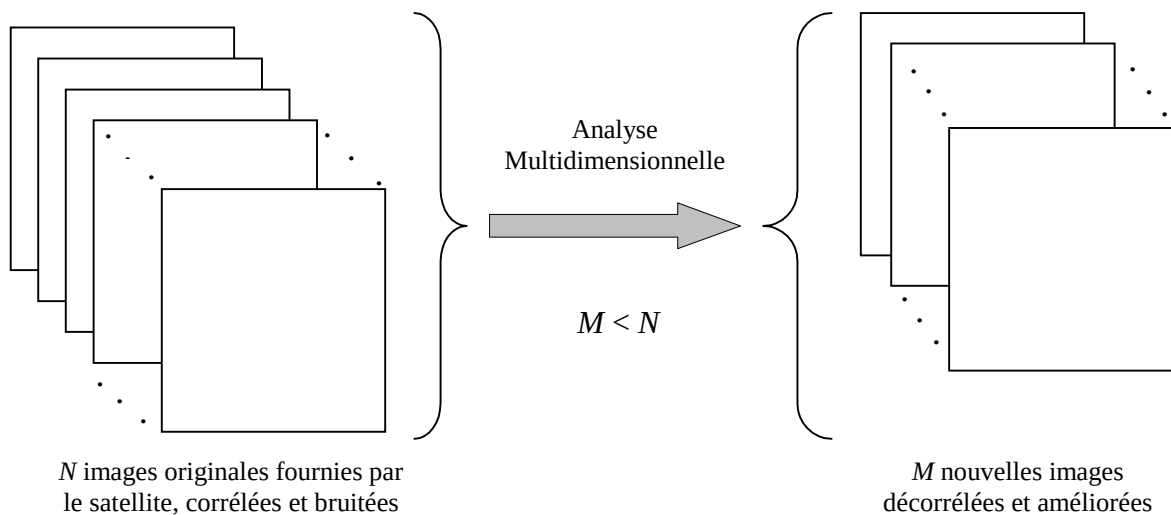


Figure 13. Analyse multidimensionnelle d'images de télédétection.

L'ACP et ses variantes sont basées sur la décomposition spectrale de la matrice de covariance des images spectrales d'entrée et celle du bruit additif pour les variantes qui font appel à cette matrice dans leur concept. Cependant, dans la pratique, la mise en œuvre de l'ACP et de ses variantes souffre de la difficulté d'une évaluation précise des vecteurs propres qui forment la matrice de la transformation linéaire. En particulier, lorsque les matrices de covariance des images d'entrée et du bruit sont mal conditionnées, pour des raisons pratiques, ou lorsque leur évaluation n'est pas précise. Ainsi, pour une application efficace de l'ACP et ses variantes, dans les conditions où l'évaluation des matrices de covariance sont imprécises, une méthode qui détermine avec précision la matrice de la

transformation linéaire, sans devoir estimer les matrices de covariance et calculer leurs décompositions spectrales, est nécessaire. Ceci peut être réalisé en utilisant les réseaux de neurones. Afin de palier aux inconvénients de l'approche statistique de l'ACP, des modèles neuronales pour l'ACP d'images multispectrales ont été proposés.

Dans cette thèse, nous avons proposé deux modèles de réseaux de neurones pour effectuer l'ACP à partir d'une image multispectrale. Les deux approches présentent les deux avantages suivants : L'ACP est directement effectuée à partir des images spectrales sans calculer la matrice de covariance de ces dernières. Le nombre d'images composantes principales à extraire n'est pas fixe comme le cas de l'ACP statistique, où le nombre d'images composantes principales dépend du nombre d'images spectrales d'entrée. De ce fait, les deux approches neuronales peuvent être appliquées pour l'analyse des images multispectrales de dimension élevée. Des études comparatives entre les approches statistiques et les approches neuronales de l'ACP et ses variantes ont été effectuées tout en faisant ressortir les avantages des approches neuronales sur les approches statistiques.

3. 3. Analyse Multidimensionnelle d'Images Radars

Les capteurs radars présentent des avantages par rapport aux capteurs optiques. Cependant, l'analyse et l'interprétation automatique des images radars sont plus compliquées que l'analyse et l'interprétation des images optiques. Les images radars possèdent des propriétés particulières qui demandent des techniques de l'analyse et de l'interprétation fondamentalement différentes de celles utilisées pour les images optiques. Dans cette thèse, la problématique du bruit multiplicatif non-Gaussien provoqué par le phénomène speckle a été soulevée tout en mettant l'accent sur l'impossibilité d'appliquer les méthodes développées pour l'analyse des images optiques sur les images radars. La réduction du speckle est donc nécessaire avant tout traitement efficace des images radars. Dans les images radars polarimétriques, en plus du problème du speckle, l'information relative à la scène observée est dispersée dans plusieurs images polarimétriques. Ces images sont corrélées ce qui pose le problème de décorrélation de ces images afin de réduire la redondance d'information et faciliter le traitement. L'autre problème posé est d'essayer d'avoir un nombre plus faible de nouvelles images qui maximisent l'information portée par les images polarimétriques. Ainsi, une représentation optimale de dimension faible (dans laquelle l'information est concentrée, les images sont décorréliées, et l'effet du speckle est réduit) devient une tâche essentielle avant toute autre tâche de traitement.

Dans ce contexte, nous avons proposé une nouvelle méthode pour l'analyse multidimensionnelle des images radars polarimétriques. La méthode est basée sur une nouvelle forme de l'ACP qui fait appel à un théorème algébrique sur le calcul matriciel. La

procédure d'implémentation de la méthode proposée est donnée en se basant sur un concept théorique prouvé. Un état d'art sur les différentes méthodes proposées dans la littérature pour l'analyse des images radars polarimétriques a été aussi donné dans le but de faire une étude comparative avec notre méthode proposée.

3. 4. Classification d'Images de Télédétection

L'utilisateur peut sur une image fournie par un satellite de télédétection civile distinguer certaines régions. En raison des faiblesses inhérentes aux différents systèmes d'acquisition, optiques et radars, cette distinction est, cependant, subjective, et d'autant plus délicate, dans le cas de la complexité des informations contenues dans les images.

La classification d'images de télédétection est une analyse quantitative qui consiste à regrouper, dans une même entité ou classe, tous les pixels ayant la même signature spectrale ou polarimétrique tout en respectant leur localisation spatiale. Le résultat est une image thématique dont le contenu ne représente plus une mesure radiométrique mais une interprétation et une catégorisation de la nature des objets terrestres associés aux pixels (Figure 14).

Cependant, le problème posé par la classification des images est la manière la plus efficace qui nous permet de traduire l'information contenue dans les images en un nombre de classes principales et faire correspondre ces classes aux différents types d'occupation du sol présents sur le site imagé. Ainsi, deux approches de méthodes de classification d'images ont été développées dans la littérature, méthodes de classification ponctuelle et méthodes de classification contextuelle.

3. 4. 1. Classification ponctuelle

Les méthodes de classification ponctuelle sont divisées en deux catégories : Méthodes supervisées et méthodes non-supervisées. Dans les méthodes supervisées, l'utilisateur intervient dès le début du processus de classification par la définition de classes *a priori*. Alors que dans les méthodes non-supervisée, l'utilisateur n'intervient qu'une fois la classification est effectuée pour interpréter le contenu des classes, sans faire appel à d'autres hypothèses sur les images ou les classes.

3. 4. 1. 1. Classification supervisée :

Les méthodes de classification supervisée peuvent être divisées en deux groupes : Méthodes basées les hypothèses statistiques, et méthodes basées sur des hypothèses géométriques.

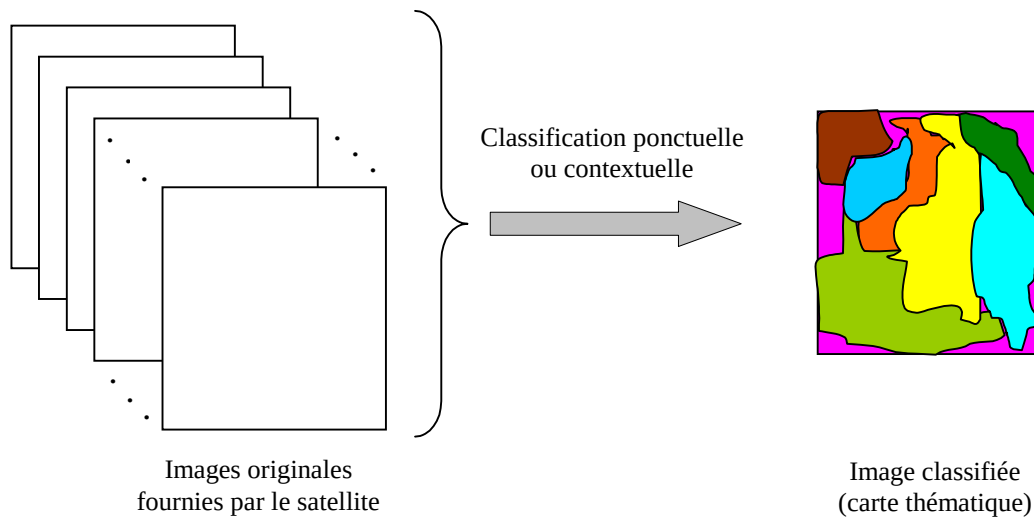


Figure 14. Classification d'images de télédétection.

Dans les méthodes statistiques, on associe à chaque classe, connue *a priori*, une fonction de densité de probabilité. Les paramètres statistiques de cette fonction sont estimées en utilisant des échantillons d'apprentissage tirés des images brutes à classifier. L'estimation des paramètres de la fonction de densité de probabilité peut être paramétrique, si la forme analytique de la fonction est connue, ou non-paramétrique si la forme exacte de la fonction est inconnue. Ces fonctions de densité de probabilité sont ensuite utilisées pour élaborer la règle de décision du classifieur supervisé.

Les méthodes géométriques font appel à des algorithmes de calcul d'indice de similarité entre un pixel quelconque et une classe connue *a priori*. Cet indice est établi en fonction d'une métrique adaptée pour chaque classe permettant ainsi de mesurer le rapprochement entre le pixel à traiter et la classe candidate. On affecte alors le pixel à la classe dont la valeur de l'indice de similarité indique le rapprochement le plus maximal.

Dans cette thèse, nous avons développé trois méthodes de classification supervisée. La première est une méthode statistique déduite de la théorie Bayésienne de décision, appelée aussi classification Bayésienne. La deuxième méthode est aussi une méthode statistique Bayésienne, mais avec l'hypothèse d'équiprobabilité des classes. Tandis que la troisième méthode est une méthode géométrique basée sur la métrique Euclidienne. Une approche de validation de la précision de classification, sans avoir recours à la réalité de terrain, est aussi donnée.

3. 4. 1. 2. Classification non-supervisée :

Dans les méthodes de classification non-supervisée, l'ensemble des échantillons d'apprentissage tiré des images brutes à classifier est vu comme un nuage de points à

découper à partir des sous-nuages disjoints correspondant aux classes recherchées. Ainsi, la classification non-supervisée peut être décomposée en deux étapes essentielles : Etape de génération de classes selon une démarche non-supervisée en utilisant les échantillons d'apprentissage, et étape de classification de l'ensemble entier des pixels selon une règle de décision établie en fonction des paramètres statistiques des classes déterminées précédemment. On obtient ainsi une image classifiée qui requiert une interprétation *a posteriori* par le thématicien. En d'autres termes, les classes obtenues par la classification non-supervisée peuvent ne pas correspondre aux classes thématiques réelles. Leur nombre peut être différent du nombre réel de classes existantes sur le terrain.

Dans la littérature on observe deux catégories de méthodes de classification non-supervisée : Méthodes statistiques non-supervisées et méthodes métriques non supervisées (ou méthodes de groupement).

3. 4. 2. 1. a- Méthodes statistiques non supervisées :

N'ayant aucune information *a priori* sur les images à classifier, le problème de classification se pose en termes d'analyse de mélange. La densité de probabilité de l'ensemble des échantillons d'apprentissage est une densité de probabilité d'un mélange. Le problème consiste donc à estimer l'ensemble des quantités probabilistes constituant la densité de probabilité de mélange. Selon les hypothèses de départ, deux approches différentes peuvent se présenter pour résoudre ce problème.

- Approche estimation : c'est une approche qui vise à estimer les quantités probabilistes. Les méthodes les plus utilisées relèvent des techniques d'estimation au sens de maximum de vraisemblances. L'algorithme le plus utilisé est l'algorithme de type EM : Estimation-Maximisation [27].
- Approche classification : Cette approche consiste à rechercher une partition de l'ensemble des échantillons d'apprentissage telle que chaque classe soit assimilable à un sous-échantillon suivant une loi de densité de probabilité. Le nombre de classes du mélange étant supposé connu. Ainsi, à partir d'une partition initiale de l'ensemble des échantillons d'apprentissage, on estime les paramètres associés à chaque classe par maximisation de la vraisemblance, on obtient ainsi une caractérisation statistique pour chaque classe. Ensuite, à partir des densités de probabilités, induites de cette caractérisation statistique, les échantillons sont classifiés. On recommence le procédé jusqu'à l'obtention de la convergence.

3. 4. 2. 1. b- Méthodes métriques non supervisées (groupement) :

Ces méthodes sont basées sur les notions de ressemblance des pixels. La ressemblance est quantifiée en mesurant la distance entre les pixels. Les procédures de mesure de

ressemblance font appel à des critères algébriques pour mesurer la qualité des groupements des pixels obtenus. Certaines minimisent un critère de dispersion inter-groupe de type quadratique, d'autres utilisent un critère calculé à partir de la matrice de covariance des groupes obtenus.

Nous citons comme exemple, la méthode des centres mobiles [28]. Elle consiste à utiliser k points pris parmi N pour définir une partition de N en k classes. La partition de l'espace de l'ensemble entier des échantillons d'apprentissage associée à k centres est un ensemble de domaines polyédriques convexes déterminés par les hyperplans médiateurs des centres. La classe est donc constituée par l'ensemble des points les plus proches de son centre que tout autre centre. On remplace alors les k points pris au hasard par les k centres de gravité de ces classes et on recommence la procédure. L'algorithme converge vers la partition optimale où la variance intra-classe est minimisée.

D'autres méthodes de classification non-supervisée plus sophistiquées ont été développées récemment dans la littérature. Dans cette thèse, nous ne considérons pas les méthodes de classification non-supervisée pour des raisons qui seront justifiées. Néanmoins, un état d'art sur les différentes méthodes de classification non-supervisée est donnée, afin de faire ressortir les avantages et les inconvénients de ces méthodes.

3. 4. 1. 3. Distinction entre méthodes statistiques et méthodes géométriques :

Bien que les méthodes géométriques supervisées et non-supervisées aient parfois des interprétations statistiques, en faisant appel aux statistiques du premier ordre et du second ordre dans la construction de la distance métrique et dans l'élaboration d'un critère d'optimisation du groupement, on ne peut proprement parler de méthodes statistiques, supervisée et non-supervisée, que lorsqu'il est fait explicitement appel à la notion de fonction de densité de probabilité.

3. 4. 2. Classification contextuelle

Dans la plupart des images de télédétection civile, les tailles des objets correspondant aux éléments au sol sont plus importantes que la taille du pixel. Par conséquent, dans une image satellitaire, deux pixels voisins sont fortement corrélés. Par exemple, sur une scène à couvert végétal, si un pixel appartient à la classe champ de blé, il est très possible que ses voisins appartiennent également à la classe champ de blé. De ce fait, un voisinage de points a plus de chance d'appartenir à la même classe. Il forme ainsi une région homogène.

3. 4. 2. 1. Définition du contexte d'un pixel :

Un pixel n'est parfaitement défini qu'une fois replacé dans son contexte [29]-[32]. Autrement dit, la caractérisation d'un pixel par sa valeur radiométrique seulement n'est pas

suffisante ; il faut aussi tenir compte de son contexte. Nous pouvons définir trois types de contextes illustrés par la figure 12 et qui sont : le contexte spectral (ou polarimétrique), le contexte spatial et le contexte temporel. Le contexte spatial est de loin le plus étudié et le plus traité dans la littérature.

3. 4. 2. 1. a- Contexte spectral/ polarimétrique:

La réflexion d'un objet au sol dans les différentes bandes spectrales ou polarimétriques est sensible à la longueur d'onde électromagnétique de chaque bande. L'analyse de l'ensemble des bandes spectrales/ polarimétriques prend donc en compte l'interaction naturelle entre les différents objets terrestres de la scène observée. Sous cette optique, les bandes spectrales/ polarimétriques peuvent être vues comme des sources d'informations différentes [33], [34]. De ce fait, le traitement conjoint des données issues des différentes sources de données relève du traitement multisources qui exploite efficacement la notion du contexte spectral/ polarimétrique des données. Généralement, cette exploitation n'est possible qu'avec le recours aux méthodes de fusion de données.

3. 4. 2. 1. b- Contexte spatial :

Le sens de la notion du contexte spatial est que la réponse spectrale/ polarimétrique et la classe de deux pixels spatialement voisins sont très liées. Autrement dit, si deux pixels sont voisins et un des deux pixels appartient à une classe quelconque, alors il y a une forte probabilité que l'autre pixel appartient aussi à la même classe. Donc, la décision de classer un pixel n'est pas basée seulement sur son contenu radiométrique, mais également sur les contenus des pixels voisins.

3. 4. 2. 1. c- Contexte temporel :

Le contexte temporel est défini pour les images multitudes représentant la même scène [35]-[37]. Ces images contiennent deux types d'informations : information correspondant aux états de surface invariants à l'échelle humaine, et information qui est variable dans l'espace des temps, par exemple l'évolution d'un champ agricole durant la saison.

3. 4. 2. 2. Méthodes de classification contextuelle :

La modélisation de l'information contextuelle permet son intégration dans la classification contextuelle d'images de télédétection. Ainsi, selon l'aspect de modélisation utilisé, nous observons dans la littérature deux familles de méthodes de classification contextuelle.

- Méthodes structurales, qui font appel aux critères texturaux dans le processus de classification.
- Méthodes statistiques, qui regroupent les méthodes de classification en utilisant les transformations et les méthodes de classification en utilisant la modélisation Markovienne des images.

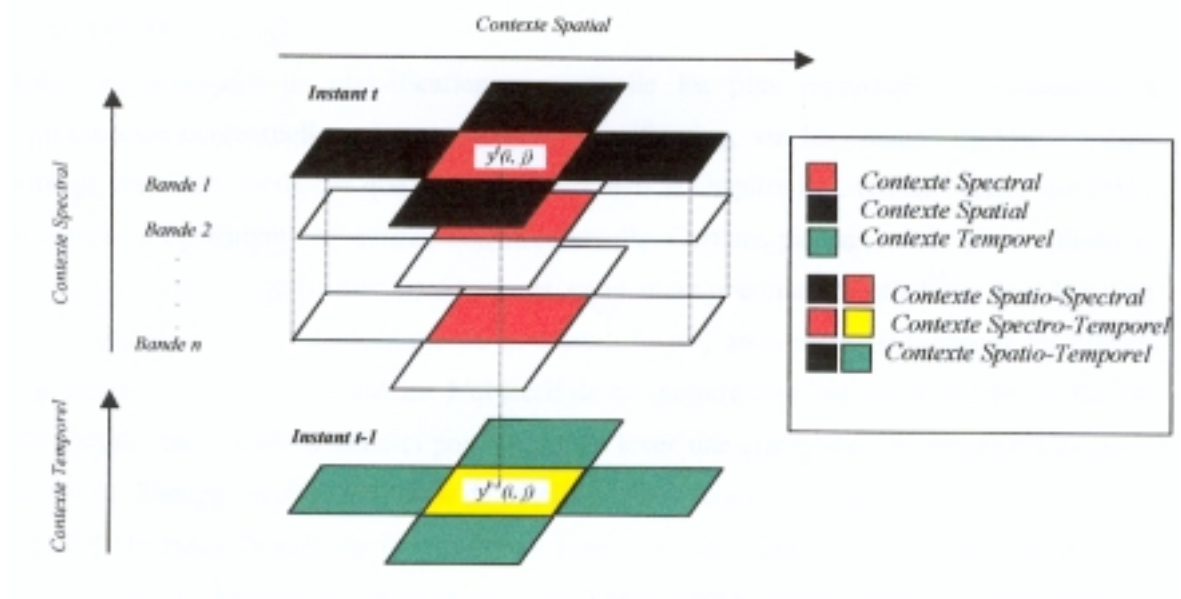


Figure 15. Les différents types de contexte d'un pixel d'image $y(i,j)$.

3. 4. 2. 2. a- Méthodes structurales :

La texture est un caractère commun à la plupart des surfaces naturelles. Bien qu'il nous soit facile de distinguer rapidement différents types de textures et d'en faire une description qualitative, proposer une définition précise de la texture n'est pas trivial.

Au sens large, la texture peut être considérée comme une mesure de la distribution spatiale des variations de tonalités [38]. Cette définition met en évidence la dépendance étroite qui lie teinte et texture. Lorsqu'une portion d'image ne présente pas de fortes variations de teinte, elle est principalement caractérisée par sa teinte (elle est faiblement texturée). En revanche, si cette portion d'image présente de fortes variations de teinte, son caractère dominant est la texture (elle est fortement texturée).

La caractérisation des textures est un thème de recherche relativement récent en classification contextuelle d'images de télédétection, par rapport aux travaux sur le contenu spectral des images. Les principales applications de la notion de texture concernent les disciplines de l'intelligence artificielle et de l'imagerie médicale. Il nous semble que ce thème reste encore peu exploité en télédétection. Il est certainement amené à se développer avec l'apparition de l'imagerie radar, jugée plus riche en informations texturales que l'imagerie optique.

Dans les méthodes de classification contextuelle par critères texturaux, appelées aussi méthodes structurales, la texture est considérée comme étant composée de motifs texturaux

élémentaires, ou primitives. Ces motifs sont fabriqués à partir de formes géométriques simples (cercle, rectangle, point, etc...). La texture est formée à partir d'une ou plusieurs primitives qui sont placées et orientées dans l'image et relativement les unes par rapport aux autres, en suivant une loi déterministe appelée loi de placement.

Ainsi, la mise en œuvre des méthodes de classification par critères texturaux nécessite la localisation des primitives de la texture, l'extraction des paramètres qui caractérisent ces primitives et enfin, l'extraction des paramètres qui caractérisent les règles de leurs dispositions spatiales. Parmi les méthodes proposées dans la littérature, nous citons : méthodes des iso-segments des niveaux de gris, méthode de la morphologie mathématique, et méthode des unités de texture. Une description détaillée de chacune de ces méthodes est donnée dans [39].

Il est bien évident que les méthodes structurales sont inadaptées à la caractérisation texturale d'une image satellitaire où il n'existe pas de primitive texturale simple et redondante. Il convient donc, dans le cas des images satellitaires, de tenir compte de caractère plus ou moins aléatoire de la texture. C'est l'objectif de la deuxième famille de méthodes, appelées méthodes statistiques, qui regroupe les méthodes de classification en utilisant les transformations et méthode de classification en utilisant la modélisation Markovienne d'images.

3. 4. 2. 2. b- Méthodes statistiques :

Ces méthodes sont basées sur la nature probabiliste de la texture. Elles visent à caractériser les lois de distribution qui gouvernent la répartition des variations de teinte sur l'ensemble ou sur une portion de l'image.

Les différentes méthodes statistiques développées dans la littérature utilisent : des calculs de variogramme [41] ; des matrices de co-occurrence [42]-[45] ; des spectres de texture [46]-[48] ; l'analyse fréquentielle par transformée de Fourier [49] ; l'analyse fractale [50]-[52] ; le modèle des champs Markoviens, le modèle Simultané Auto-Régressif (SAR) [53]-[55], etc..... Cependant, toutes ces méthodes peuvent être divisées en deux classes de méthodes : méthodes en utilisant les transformations et méthodes en utilisant la modélisation Markovienne. Nous citons dans ce qui suit les plus utilisées dans littérature.

- Méthodes utilisant les transformations :

Plusieurs méthodes de classification contextuelle, en utilisant les transformations, ont été développées dans la littérature. Ainsi, on observe diverses méthodes de mesure des propriétés locales de la texture. Ces méthodes se basent soit sur les caractéristiques locales du pixel en fonction de ses pixels voisins, méthodes appelées transformations locales, soit

sur la génération de matrices de dépendance de niveaux de gris, puis la détermination des paramètres texturaux à partir de ces matrices [61], [40].

Le principe général des méthodes des transformations locales est le suivant : considérons la fonction $I(i,j)$ représentant l'intensité du pixel ayant pour coordonnées (i,j) . On définit la transformée texturale F qui associée à chaque pixel $I(i,j)$ un nouveau pixel dont l'intensité $F\{I(i,j)\}$ est la mesure de la distribution du niveau de gris des pixels qui l'entourent. Elle représente une quantification des propriétés statistiques locales de ce pixel. Cette quantification est une modélisation du contexte spectral-spatial du pixel. Parmi les propriétés locales de la texture, nous citons par exemple, la moyenne, la variance, la dissymétrie, l'aplatissement, la dynamique, l'énergie, etc.... La variance est de loin, la propriété locale de texture la plus utilisée. Elle mesure la dispersion du pixel central par rapport à la moyenne de son voisinage spatial. Elle a été utilisée comme indice de texture et critère de discrimination entre les classes confuses [30], [58], [62].

Une autre méthode proposée pour la classification contextuelle des images multispectrales en utilisant les Matrices de Co-occurrences de Niveaux de Gris (MCNG) [40]. La méthode fait appel aussi aux critères statistiques dans la règle de décision. Ainsi, le contexte spectral-spatial est modélisé par un indice de texture qui est intégré dans la classification ponctuelle via un terme pondéré par rapport à l'information spectrale. Cependant, l'inconvénient major des méthodes de transformation locale, en général, est le problème du choix adéquat de la taille du voisinage optimal. Une étude détaillée du problème de l'échelle et de la texture dans la classification des images numériques est donnée dans [61]. Une large fenêtre d'analyse nécessaire pour une stabilité de la mesure de texture entraîne une dégradation des contours et des structures linéaires. Inversement, une petite taille de la fenêtre d'analyse minimise l'effet des contours mais n'arrive pas à produire une mesure de texture stable.

Le principe des autres méthodes utilisant les transformations est de décomposer l'image à classifier sur une base vectorielle d'un nouveau système de coordonnées, qui doit être lié aux fréquences spatiales. Parmi les transformations qui permettent l'analyse de texture nous citons : la transformée de Fourier ; la transformée en cosinus, et la transformée en ondelettes. Les coefficients de ces transformées représentent la décomposition de l'image originale en termes de composantes de fréquences spatiales qui expriment les transitions des niveaux de gris dans le domaine spatial. La texture d'une image peut être alors analysée par l'étude des coefficients ou du spectre de puissance de ces transformées.

La transformée de Fourier ne permet pas de modéliser que les phénomènes stationnaires et n'offre qu'une analyse globale de la texture sur une seule échelle. Or, la texture est phénomène non stationnaire [56], très bien localisé en temps et sa représentation est de

plus en plus précise en passant des échelles grossières vers les échelles fines [57]. Il est donc intéressant de définir et d'utiliser d'autres méthodes prenant en compte la résolution fréquentielle relative à une analyse multi-échelle et la résolution temporelle relative à une analyse temporelle locale. La transformée en ondelettes associée à l'analyse multi-résolution, est l'une de ces méthodes.

L'apport de la transformation en ondelettes en information de texture via les coefficients d'ondelettes a déjà été étudié [58]-[60], [56]. Les coefficients d'ondelettes générés par la transformée en ondelettes d'une image décrivant la différence d'informations entre deux approximations successives à deux résolutions successives de cette image. Ils modélisent parfaitement les structures et les détails invisibles sur les petites échelles. Le contexte spectral-spatial est alors bien mis en évidence. Le principe est de considérer les coefficients d'ondelettes comme image supplémentaire dans la classification des images multispectrales/ polarimétriques. Ainsi, ces coefficients sont utilisés pour extraire les indices de textures, en l'occurrence les variances locales, et l'image résultat est intégrée dans la classification ponctuelle. De ce fait, les structures qui sont inaccessibles à la résolution courante de l'image sont intégrées dans la classification. Le choix de la résolution (ou échelle) à laquelle on extrait les structures à exploiter dépend essentiellement de la nature des thèmes dont on veut augmenter la discrimination par rapport aux autres thèmes. La transformée de Gabor qui fait partie de la même famille de transformée que la transformée d'ondelettes a déjà été utilisée comme indice de texture dans la classification d'images satellitaires [30].

- Méthodes utilisant la modélisation Markovienne d'images :

Le formalisme des champs aléatoires de Markov dispose d'une grande capacité pour l'incorporation séparée ou conjointe du contexte spatial-temporel [31]. Les pixels d'une image ne sont pas indépendants les uns des autres. La modélisation de l'image par un champ de Markov va cependant permettre de restreindre cette dépendance aux pixels voisins uniquement. A travers un champ de Markov, le contexte spatial est modélisé par une fonction d'énergie et le contexte temporel par une matrice de transition des classes entre deux instants différents.

Il est hors de notre propos de faire un bilan exhaustif des mérites et défauts de toutes les méthodes de classification contextuelle ainsi développées dans la littérature. Notre objectif était de mettre au point des procédures de classification contextuelle des images satellitaires à partir des critères texturaux. Nous avons été amenés à faire un choix parmi toutes ces méthodes. Ce choix a été influencé par la simplicité de mise en œuvre des méthodes, tant du point de vue de la compréhension des concepts que du point de vue de l'implémentation, par leur relative efficacité reconnue et par leur originalité. Il s'agit de la classification contextuelle en utilisant la modélisation Markovienne.

Dans cette thèse, nous avons consacré un chapitre essentiellement aux méthodes de classification contextuelle basées sur la modélisation de l'information contextuelle spatiale par les champs aléatoires de Markov "*Markov Random Field (MRF)*". L'approche Bayésienne associée à la modélisation Markovienne présente une fonction d'énergie non-convexe dont le minimum est la caractéristique de l'image classifiée recherchée. La solution d'un tel problème vient des méthodes d'optimisation. Ainsi, nous proposons dans ce chapitre l'étude et le développement de plusieurs méthodes d'optimisation stochastiques et déterministes.

3. 5. Fusion d'Images en Télédétection

Les images acquises sur une même scène par différents capteurs sont partiellement redondantes puisqu'elles représentent la même scène, et partiellement complémentaires, puisque les capteurs ont des caractéristiques différentes, et que les mécanismes d'interaction de l'onde électromagnétique et de la surface imagée diffèrent. Autrement dit, l'examen de chacune des images d'une même scène, fournies par les différents capteurs, montre qu'il existe des informations significatives propres à chacun des capteurs qui peuvent être détectées facilement dans une image et difficilement dans une autre image. Il montre aussi qu'il y a une certaine relation existant entre les images des différents capteurs. En outre, l'information fournie par chaque capteur est incomplète et imprécise. Il est intéressant alors d'une part de traiter la redondance entre les différents capteurs pour réduire l'imprécision contenue dans les données de chaque capteur, et d'autre part d'utiliser leur complémentarité, en exploitant les informations significatives portées par chacun des capteurs, pour obtenir une interprétation qui ne pourrait être obtenue à partir des données d'un seul capteur. Ceci nécessite l'exploitation simultanée de l'ensemble des images fournies par les différents capteurs.

Le problème qui se pose réside dans la recherche d'une manière efficace à l'exploitation simultanée de ces différentes sources d'informations. Il s'agit ici d'élaborer un processus appelé fusion de données multisources. La stratégie à suivre dans la fusion devrait dépendre de l'objectif visé et de la manière dont la fusion sera effectuée. Ainsi, une interprétation plus cohérente de la scène imagée, dans laquelle l'incertitude associée est diminuée, peut être obtenue.

Cependant, le problème de définition exacte du terme de fusion de données, les ambiguïtés dans la signification ainsi que l'objectif inhérent à l'utilisation de la fusion de données (souvent rencontrées dans la littérature) devraient être mises en évidence. En fait, le terme fusion de données, quoique récent, désigne une démarche adoptée spontanément dans de nombreux domaines.

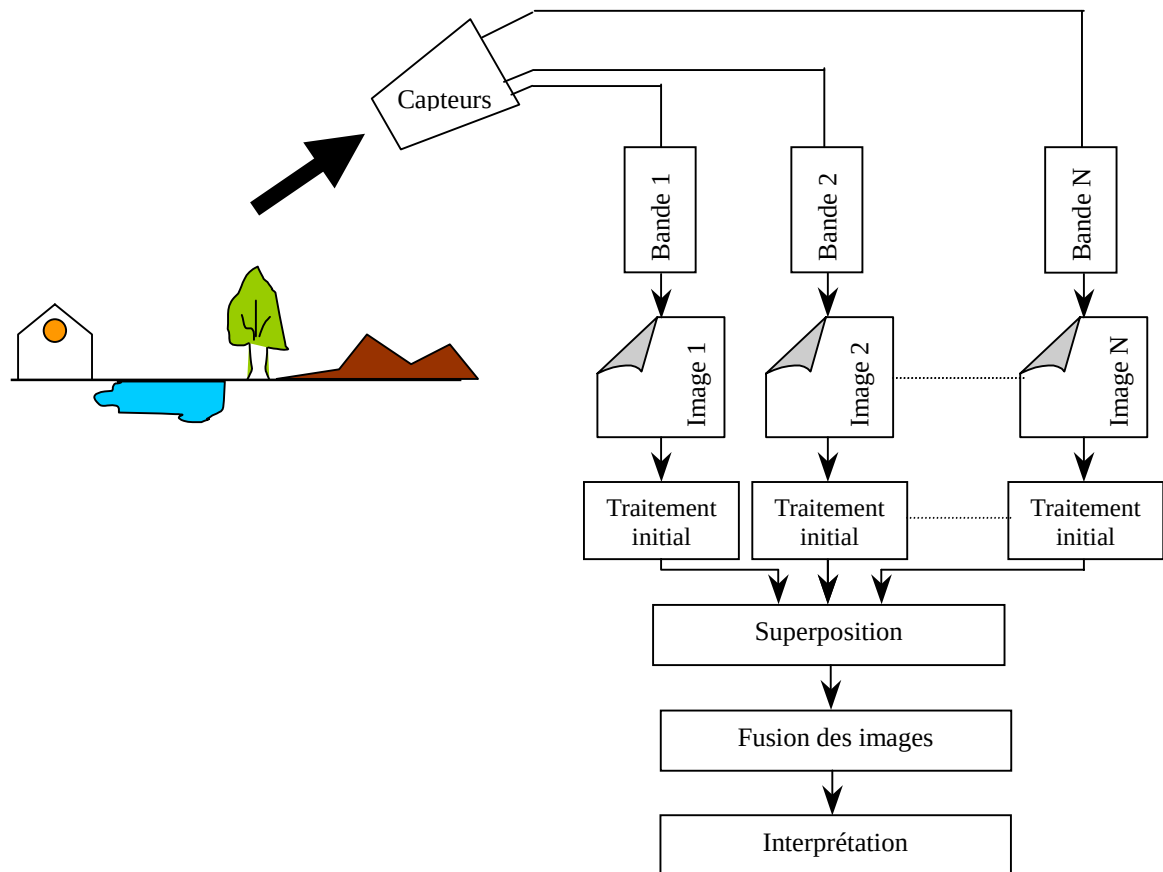


Figure 16. Processus de traitement et de fusion pour l'interprétation des images multisources.

D'après Wald [33] : «La fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses. Elle vise à l'obtention d'informations de plus grande qualité ; la définition exacte de « plus grande qualité », dépendra de l'application». D'un point de vue général, la présentation d'une image de qualité requiert au préalable l'accomplissement d'un ensemble de tâches élémentaires distinctes (Figure 13) : acquisition de l'image dans une bande spectrale particulière, traitement initial, superposition des images, fusion des images puis interprétation. La finalité de la fusion des images multisources est de rendre l'interprétation plus aisée.

Selon les objectifs fixés au préalable, la fusion des images peut s'opérer à différents niveaux pour lesquels nous pouvons distinguer trois approches : fusion bas niveau ou de pixels [63], fusion d'attributs ou de primitives [58] et fusion haut niveau ou de décisions [31]. Un autre type de fusion d'informations, récemment proposée dans la communauté d'intelligence artificielle, est la fusion des classifieurs. Dans cette thèse, nous avons proposé trois méthodes de fusion d'imagerie de télédétection. Le but est d'améliorer la classification de la scène imagée, en considérant la fusion comme une deuxième

classification qui succède à celle des classifieurs. Dans ce qui suit, les trois approches de fusion des images sont exposées. Ensuite la fusion de classifieurs est donnée.

3. 5. 1. Fusion bas niveau

La fusion bas niveau s'opère directement sur les pixels de l'image; ils constituent les mesures les plus proches de la réalité physique de l'objet observé. Les pixels à fusionner ont généralement subi quelques traitements liés principalement aux conditions de prise de vue. Ce type de fusion a été appliqué avec succès en télédétection, en particulier, la fusion d'images multispectrales de résolution spatiales différentes. Le principe est de simuler des images de haute résolution spectrale à la meilleure résolution spatiale disponible. Par exemple, dans le cas de l'image SPOT, il est désirable à partir des trois images spectrales de 20m de résolution spatiale et de l'image panchromatique de 10m de résolution spatiale d'avoir trois nouvelles images spectrales 10m de résolution spatiale. Cette manière de fusion traite la radiométrie des images spectrales, de ce fait, la fusion est opérée au niveau du pixel. Dans ce contexte, deux catégories différentes d'approches ont été développées dans la littérature.

La première approche de la première catégorie est l'approche de la combinaison relative des luminances des différentes bandes (à haute et à faible résolution) en exploitant l'information contextuelle spectrale du pixel. Les images spectrales synthétiques à la résolution haute sont obtenues par la transformée de Brovey [75]. La deuxième approche est la transformée IHS "*Intensity Hue Saturation*" dont le principe est la projection des images originales à faible résolution dans une nouvelle base, ensuite la substitution de l'une des nouvelles composantes par l'image à haute résolution et à la fin pratiquer la projection inverse vers la base d'origine [76], [77]. Cependant, ces deux approches nécessitent que les images spectrales et panchromatique soient géocodées et superposables, c.-à-d. représentent la même zone géographique (recalage géométrique) et de même taille.

Dans la deuxième catégorie, nous citons l'approche basée sur l'analyse multirésolution par l'algorithme de Mallat et la transformée en ondelettes. L'approche intègre le contexte spectro-spatial et a été élaborée avec l'ambition de préserver l'information spectrale des images dont on veut améliorer la résolution spatiale. Elle se propose donc, comme une solution aux problèmes que présentent la transformée de Brovey et la transformée IHS. Elle est connue sous le nom ARSIS (Amélioration de la Résolution Spatiale par Injection de Structures). Ainsi, comme son nom l'indique, l'approche consiste à extraire l'information spatiale (les structures) contenue dans l'image de meilleure résolution pour l'injecter dans les images de faible résolution. L'approche a été appliquée avec succès pour la fusion d'images Landsat-TM et SPOT panchromatique [58]. L'approche, cependant, s'appuie sur des hypothèses qui ne sont pas toujours vraies, à savoir : L'écart temporel

entre les deux groupes d'images à fusionner n'influe pas sur l'aspect des structures observées ; l'existence d'une similitude entre les structures observées dans les différentes bandes spectrales.

3. 5. 2. Fusion de primitives

L'examen des images l'une après l'autre est une tâche fastidieuse pour le photo-interprète. En effet, celui-ci n'a pas la capacité d'intégrer mentalement toutes les informations provenant de chacune des images multisources. C'est pourquoi, une procédure de fusion des différentes sources d'images devient nécessaire afin de générer une nouvelle image composite dont le contenu informatif a été amélioré. Ceci est possible grâce à la fusion d'attributs ou de primitives qui nécessite l'extraction de primitives, à partir de chaque image source, caractérisées par des contours, des régions (segmentation), des textures ou composantes fréquentielles [64]. Des critères de fusion complexes sont alors appliqués sur ces primitives en tenant compte de la forme, de l'orientation et du voisinage [65]. L'objectif de la fusion consiste à maximiser l'information significative dans une seule image afin de faciliter l'interprétation.

La fusion de primitives des deux images, schématisée dans la figure 14, consiste à représenter chaque image à fusionner par un ensemble d'informations significatives, appelées primitives, extraites des deux images, puis combiner ces informations au moyen d'une règle de sélection afin de générer un ensemble d'informations significatives (primitives importantes) pour l'image composite. Celle-ci est reconstruite à partir de cet ensemble d'informations. Ainsi, l'extraction des primitives constitue l'étape la plus importante dans la caractérisation de l'information significative. Pour cela, de nombreuses méthodes ont été mises en œuvre pour extraire ces primitives.

Les techniques de fusion de primitives des deux images se distinguent principalement par la manière dont sont extraites les primitives. La décomposition hiérarchique constitue l'approche la plus usitée pour l'extraction des primitives [67]. A chaque niveau de pyramide, les primitives des deux images sont fusionnées, au moyen d'une règle, pour générer les primitives de l'image fusionnée. La reconstruction, à partir de la pyramide inverse, permet de produire l'image fusionnée. Une approche similaire a été proposée dans [68]-[70] qui consiste à fusionner des images thermique et visuelle. Les primitives de l'image sont caractérisées, cette fois ci, par le contraste défini comme étant le rapport de l'image avant filtrage et l'image interpolée. Cela va permettre de générer une pyramide de contraste pour chacune des images à fusionner. L'utilisation de la décomposition en ondelettes constitue un autre outil récent et plus performant pour décrire les primitives de l'image [71]. Celles-ci sont caractérisées par des coefficients dits d'ondelettes. Des règles de fusion adaptées dans le plan des ondelettes sont mises en œuvre pour améliorer la qualité de l'image fusionnée [72]-[74].

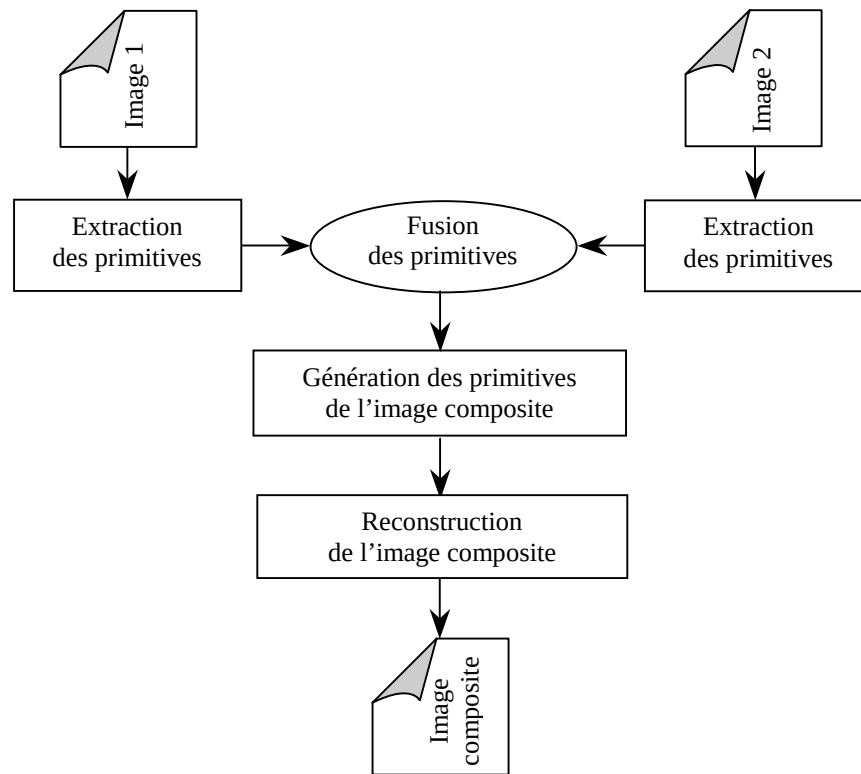


Figure 17. Fusion de primitives.

3. 5. 3. Fusion haut niveau

Afin de caractériser complètement un objet terrestre (classe), il est nécessaire d'exploiter simultanément les images multisources disponibles au moyen d'une synergie adéquate afin de mieux faire ressortir les informations les plus significatives et mettre à profit la complémentarité des images disponibles. Cette synergie doit permettre la génération d'un ou plusieurs nouveaux paramètres ou fonctions caractéristiques au moyen d'un processus de fusion qui doit être simple, mais suffisamment efficace pour accroître la fidélité de la caractérisation de l'objet terrestre.

Cet objectif peut être atteint par l'élaboration de stratégies de fusion de haut niveau aptes à comprendre la nature, la qualité et la pertinence de l'information recherchée. Cette fusion haut niveau est une sorte d'extension de la fusion de primitives. En effet, les primitives, représentées ici par les paramètres ou les fonctions caractéristiques estimées des images à fusionner, sont d'abord classées puis fusionnées au moyen d'un modèle décisionnel pour affiner la classification. Ce modèle est conçu en utilisant la théorie Bayésienne [31], la théorie des ensembles flous [66], [87], [88], ou la théorie de l'évidence [89] afin de résoudre les problèmes liés à l'imperfection de l'extraction des primitives lors de la classification d'images.

3. 5. 4. Fusion de classifieurs

L'idée principale de la fusion de classifieurs est l'augmentation de performances, autrement dit amélioration de la classification de la scène imagée. Cette augmentation de performance peut être plus de fiabilité dans les réponses, ou moins de rejet, ou bien les deux en même temps. En considérant la fusion des classifieurs comme une deuxième classification qui succède à celles des classifieurs, le principe est de fusionner les réponses des classifieurs sans tenir compte de leurs caractéristiques internes. Les réponses fournies par les classifieurs peuvent être divisées en quatre types :

- ❑ **Type Classe.** La réponse du classifieur est une classe. Il considère que la forme d'entrée (un pixel dans notre cas) appartient à cette classe.
- ❑ **Type Ensemble.** La réponse du classifieur est un ensemble de classes. Il considère que la forme d'entrée appartient à une des classes de cet ensemble, mais il ne donne aucune indication supplémentaire qui pourrait nous aider à discriminer parmi ces classes.
- ❑ **Type Rang.** La réponse est un classement sur toutes les classes ou seulement sur une partie des classes. La classe mise au premier rang du classement est la classe du dernier rang est la classe la moins probable. Le classifieur indique ce classement en donnant comme sortie un vecteur des rangs.
- ❑ **Type Mesure.** Le classifieur donne, pour toutes les classes, des mesures qui indiquent en quelque sorte, directement ou indirectement, à quel degré le classifieur considère chaque classe comme la vraie classe de la forme d'entrée. La sortie du classifieur est donc un vecteur de mesures.

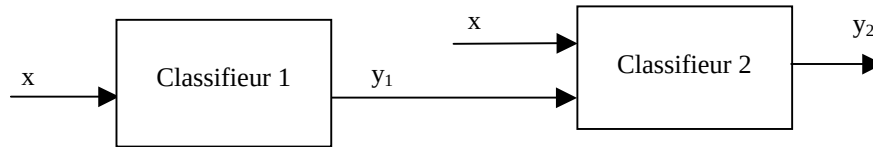
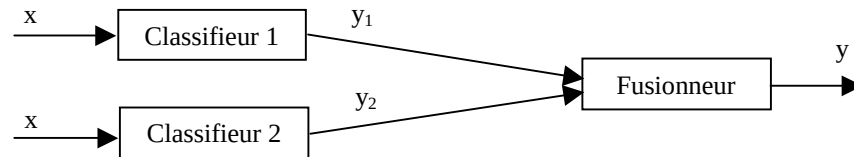
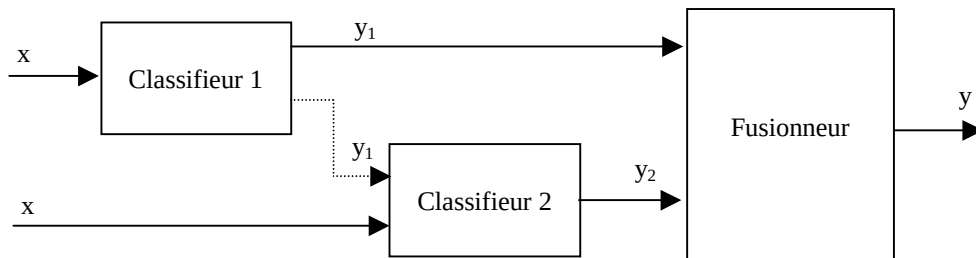
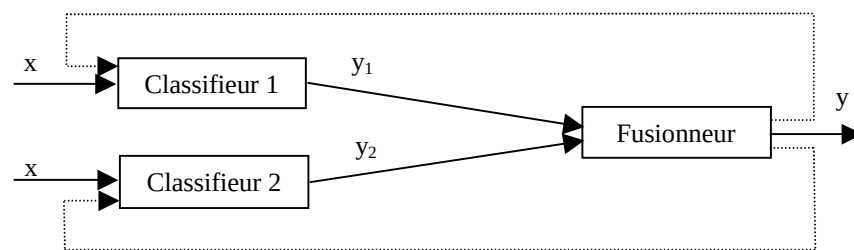
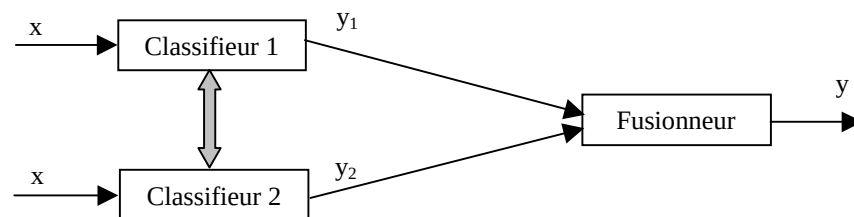
Il est évident que les classifieurs de type rang ou de type mesure peuvent être facilement convertis en type classe. Pour le type rang, il suffit de choisir la classe qui est au premier rang. Les autres classes ne seront pas utilisées. Pour le type mesure, il suffit de choisir la classe qui donne la meilleure mesure (mesure minimum ou maximum, selon que la mesure est une mesure croissante ou décroissante). Pour le type ensemble, la conversion est moins évidente. Une manière de la réaliser est de choisir le rejet quand l'ensemble proposé par le classifieur contient plus d'une classe, et de choisir la classe proposée si l'ensemble ne contient qu'une seule classe. Différentes méthodes de fusion de classifieurs ont été proposées dans la littérature qui peuvent être regroupées selon les critères suivants.

- ❖ **Ordre d'exécution des classifieurs.** Dans cette catégorie de méthodes, on observe deux types de fusion : fusion séquentielle (figure 18), fusion parallèle (figure 19) et

fusion série-parallèle (figure 20). Dans la fusion séquentielle, un ou plusieurs classifieurs sont utilisés pour modifier l'exécution d'autres classifieurs. Cette modification peut consister dans l'agencement des autres classifieurs ou dans les données sur lesquelles ils travaillent. L'ordre d'exécution des classifieurs est important et s'il est changé, le résultat final peut être différent. Quelques exemples de la fusion séquentielle peuvent être trouvés dans [92]. Dans la fusion parallèle, les résultats obtenus par les classifieurs n'affectent pas l'exécution des autres classifieurs. Par conséquent, l'ordre d'exécution des classifieurs n'importe pas. Ils peuvent être exécutés tous en même temps ou les uns après les autres. La fusion série-parallèle est un mélange de la fusion séquentielle et fusion parallèle.

- ❖ **Fusion avec bouclage.** Dans la fusion avec bouclage (figure 21), le résultat de la fusion peut à son tour être utilisé par les classifieurs et modifier ainsi leurs sorties. Ce schéma de fusion est très puissant puisque le système entier (les classifieurs et le fusionneur) devient un système dynamique qui peut être beaucoup plus performant qu'un système statique. En contre partie, les systèmes dynamiques non linéaires sont beaucoup plus difficiles à analyser et à contrôler.
- ❖ **Fusion avec interaction entre les classifieurs.** Dans ce type de fusion (figure 22), Les classifieurs interagissent entre eux, et modifient éventuellement leurs résultats avant que la fusion soit réalisée.
- ❖ **Fusion utilisant directement l'entrée.** Dans ces méthodes de fusion, les données d'entrée des classifieurs sont utilisées en même temps que les réponses des classifieurs. Selon que l'algorithme de fusion a accès à l'entrée ou pas, deux cas de figures se présentent : soit la sortie du fusionneur est en fonction de seulement des sorties des classifieurs ; soit elle est en fonction des sorties des classifieurs et de l'entrée (figure 23).

La fusion de classifieurs de types différents peut se faire avec trois méthodes différentes [93]. La première consiste à convertir la sortie fournie par chaque classifieur en le type le plus simple parmi les classifieurs, et faire la fusion selon ce type. Cette méthode entraîne une perte d'informations des sorties des classifieurs. La deuxième méthode convertit la sortie fournie par chaque classifieur en le type le plus élevé parmi les classifieurs, et faire la combinaison selon ce type. Cette méthode n'entraîne pas de perte d'informations des sorties des classifieurs, mais entraîne souvent une perte de performance de fusion, du fait que la méthode utilisée n'est pas adaptée pour des sorties converties des autres types. Enfin, une façon de fusionner des classifieurs de types différents est d'utiliser une méthode de fusion qui est capable de fusionner directement des sorties de types différents.

**Figure 18.** Fusion séquentielle de classifieurs.**Figure 19.** Fusion parallèle de classifieurs.**Figure 20.** Fusion série-parallèle de classifieurs.**Figure 21.** Fusion avec bouclage de classifieurs.**Figure 22.** Fusion avec interaction entre les classifieurs.

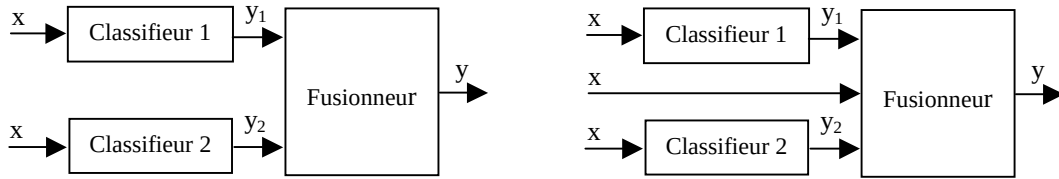


Figure 23. Fusion de classifieurs utilisant directement les entrées.

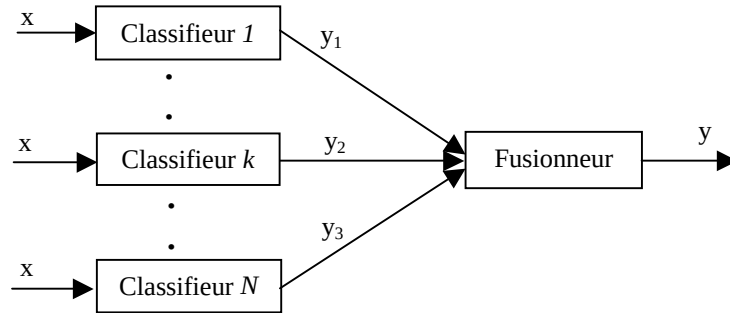


Figure 24. Cadre de travail choisi.

Dans cette thèse, nous avons étudié et implémenté les méthodes de fusion des classifieurs de type classe appliquées à la classification des images de télédétection. En supposant que les classifieurs ne changent pas, le fusionneur est optimisé via le concept théorique de la méthode de fusion. Ainsi, nous avons choisi un cadre de travail bien précis, caractérisé par la figure 24, qui est le suivant :

- Le schéma de fusion sera parallèle, dans lequel il y a plusieurs classifieurs en parallèles et un fusionneur.
- Il n'y a pas d'interaction entre les classifieurs.
- Il n'y a pas de rebouclage du fusionneur vers les classifieurs.
- Les classifieurs sont fixes et ne changent pas.
- Les classifieurs peuvent utiliser les mêmes données, des données différentes ou des données partiellement différentes.
- Le fusionneur n'a pas accès aux entrées des classifieurs.

3. 6. Filtrage, Détection de Contour et Segmentation d'Images

Les méthodes de filtrage, détection de contour et segmentation d'images ne seront pas considérées dans cette thèse, mais le lecteur peut se référer aux références citées dans la bibliographie pour plus de détails.

3. 6. 1. Filtrage d'images

Les méthodes de filtrage des images optiques et radars ont pour but d'atténuer l'effet du bruit additif ou du speckle dans l'image. Les méthodes de filtrage du speckle dans les images radars se subdivisent en deux familles : les méthodes heuristiques et les méthodes initiées [94], [26].

Les méthodes heuristiques permettent de réduire le bruit additif/ ou le speckle sans tenir compte de son caractère additif/ ou multiplicatif et sans connaissance *a priori* de ses lois de probabilités. Elles sont fondées uniquement sur le critère de voisinage. Le filtre moyenne en est un exemple qui permet une réduction efficace dans les zones homogènes. Il associe à chaque pixel la moyenne de ses voisins. On peut utiliser plusieurs types de moyenne, en pondérant différemment les voisins. Toutefois, son défaut majeur provient du fait que les contours et les textures sont trop lissés donnant une impression de flou. Par contre le filtre médian préserve correctement les structures significatives mais au détriment d'un filtrage moins efficace nécessitant un traitement itératif avec une fenêtre de petite taille. Le filtre géométrique (ou morphologique) repose sur le principe de la modification de la forme de la surface des niveaux de gris. La valeur estimée du pixel est une combinaison non linéaire des pixels voisins. Son atout réside dans la réduction efficace du speckle dans les régions homogènes par l'application de plusieurs itérations. Cependant, il dégrade la texture et les réponses ponctuelles. Le filtre sigma est fondé sur la distribution de Gauss dont l'information utile (pixel estimé sans bruit) est contenue à l'intérieur de l'intervalle trois sigma (où sigma est l'écart type de la distribution gaussienne). En dehors de cet intervalle, le pixel traité sera considéré comme un bruit ; il sera donc supprimé. Ses performances sont proches de celui du filtre médian.

Les méthodes initiées sont destinées pour le filtrage du speckle. Elles sont fondées sur le caractère multiplicatif du speckle. La mise en œuvre de filtres requiert sa modélisation et une connaissance *a priori* de ses lois de probabilité. Pour cela, différents filtres adaptatifs ont été développés dont les plus connus sont ceux de Frost, Kuan et Lee. Ceux-ci sont des filtres adaptatifs qui utilisent les statistiques locales de l'image calculées sur une fenêtre carrée. Un filtrage efficace est généralement obtenu lorsque la fenêtre considérée correspond à une zone homogène de l'image. Par contre, lorsque cette zone renferme des singularités, les performances de ces filtres se trouvent fortement dégradées. Cela provient du fait que l'hypothèse d'un speckle totalement développé n'est plus satisfaite. C'est pourquoi, plusieurs variantes de ces filtres ont été développées pour améliorer le filtrage dans les situations défavorables. Elles reposent essentiellement sur une mesure locale qui permet la vérification de l'existence d'une zone homogène, zone hétérogène afin d'adapter une procédure particulière de filtrage.

3. 6. 2. détection de contours

Les contours d'une image renferment une information très riche sous une forme très condensée. Il existe une grande variété de contours. En fait, un contour est une zone de transition séparant deux plages homogènes qui présentent des caractéristiques différentes (niveau de gris, texture, etc....). Lorsqu'on veut extraire automatiquement les contours des éléments d'une image, on utilise une fonction qui explore les divers niveaux de gris des pixels d'un masque. Celui-ci se déplace sur la plus nette en direction et en valeur. Ceci permet de détecter les pixels pour lesquels le changement d'intensité correspond précisément aux contours de l'image.

Les méthodes classiques de détection de contours procèdent essentiellement par l'application d'un opérateur local de différentiation précédé d'un opérateur de filtrage. Ces méthodes se divisent en deux familles des approches : Approche du gradient qui consiste à calculer le gradient et l'extraction des extrema locaux de la norme dans la direction du gradient ; approche du Laplacien qui consiste à calculer Laplacien et détermination de ses passages par zéro. Chacune de ses deux familles pouvant être subdivisée en deux classes : Algorithmes à caractéristiques fréquentielles figées dits du premier niveau telles que les masques de Priwitt, Sobel, Kirsh ; algorithmes à caractéristiques fréquentielles non figées dits du second niveau et regroupent les détecteurs de contours optimaux selon les critères de Canny (bonne détection, bonne localisation sous la contrainte de la réponse unique). Ces techniques peuvent être regroupées suivant le schéma ci-dessous (figure 25).

3. 6. 3. Segmentation d'image

La segmentation d'image a pour but de partitionner l'image en un ensemble correspondant à des objets ou à des parties significatives d'objets de la scène observée. Les méthodes de segmentation développées dans la littérature diffèrent d'une part par le type d'indice visuel qu'elles utilisent : les primitives peuvent être classées selon leurs dimensions géométriques comme les points remarquables, les contours, les régions, les structures. Elles peuvent être aussi spécifiées de manière plus sémantique, comme les éléments de texture.

Il existe deux approches de segmentation, une première approche basée sur la détection des contours qui implique la recherche des discontinuités locales de la fonction niveau de gris de l'image. La seconde approche fondée sur l'extraction de régions qui nécessite quant à elle la détermination des zones homogènes (en niveau de gris, même texture, même objet rigide en mouvement). Dans le cas d'images réelles par exemple, les contours correspondent aux frontières d'objets et les régions à leur surface.

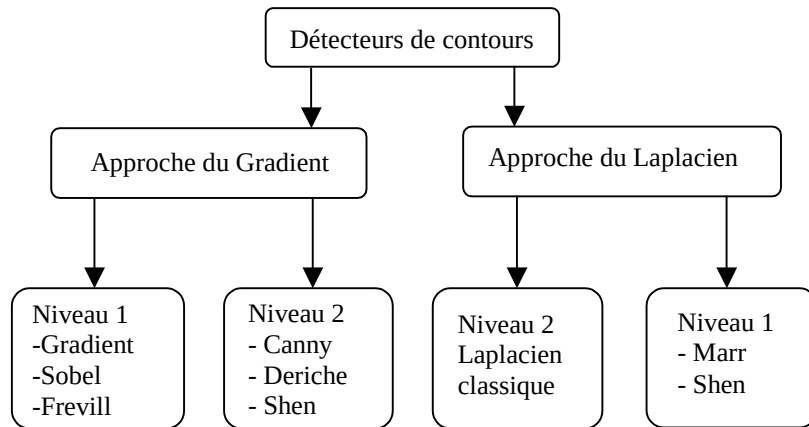


Figure 25. Techniques de détection de contours.

Ces deux approches, contours et régions, sont duales en ce sens qu'une région définit une ligne par son contour et qu'une ligne fermée définit une région. Elles conduisent cependant à des algorithmes complètement différents et n'aboutissent pas aux mêmes résultats. La segmentation par région consiste à grouper les points adjacents dont les attributs varient d'une façon négligeable. Ce type de segmentation peut utiliser le seuillage qui est une des méthodes les plus simples, consistant à affecter à l'objet ou au fond de la scène le pixel ayant un niveau de gris supérieur ou inférieur au seuil. Cette méthode est surtout utile lorsque les objets de la scène sont placés sur un fond uniforme. En outre, il existe d'autres méthodes telles que : La division qui consiste à diviser une région chaque fois qu'un critère d'uniformité sur cette région est faux ; la fusion, inversement à la division, fusionne deux régions connexes à partir d'un découpage initial de l'image, dès qu'un critère d'uniformité est vérifié. La segmentation basée sur la détection de contours consiste à calculer la variation d'intensité lumineuse en chaque pixel de l'image. On choisit comme points de contours les points pour lesquels le gradient est supérieur à un seuil ou en recherchant les passages par zéro du Laplacien. D'autres méthodes plus récentes sont basées sur l'apport des analyses multirésolution et multiéchelle à la segmentation d'images optiques et radars ont été proposé [94]. Ces méthodes utilisent les coefficients d'ondelettes. Les résultats obtenus semblent satisfaisants est promoteurs.

4. Conclusion

Dans ce chapitre nous avons présenté les chaînes de formation des images de télédétection spatiale civile, optiques et radars. Nous avons décrit les différentes distorsions que les images ont subies durant le processus de formation, en faisant ressortir les faiblesses inhérentes aux différents systèmes d'acquisition, optiques et radars, et la nature des images

obtenues. Ensuite, une méthodologie globale de traitement d'images de télédétection est proposée. Cette méthodologie fait appel à plusieurs méthodes de traitement d'images. Un état de l'art de chaque méthode est donné. Ces méthodes seront détaillées dans les chapitres suivants. Ainsi, nous évoquons, dans l'ordre dans les prochains chapitres, la modélisation, l'analyse, la classification, et la fusion des images de télédétection spatiale civile optiques et radars.

Références

- [1] S. D. Drury, *Image Interpretation in Geology*. Allen & Unwin, London, 1987.
- [2] R. Escadafal, *Caractérisation de la Surface des Sols Arides par Observation de Terrain et par Télédétection*. Ouvrage objet d'une Thèse de Doctorat d'Etat en Pédologie, Edition ORSTOM, Collection Etude et Thèses, Paris, 1989.
- [3] M. C. Girard & C. M. Girard, *Télédétection Appliquée Zones Tempérées et Intertropicales*. Edition Masson, Paris, 1989.
- [4] F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications*, Artech House, 1990.
- [5] S. Chitroub, *Implémentation d'un Processus d'Analyse et de Classification d'Images Satellitaires. Application à la Caractérisation des Etats de Surface*, Thèse de Magister, Institut d'Electronique, U.S.T.H.B, Alger, 03/94-M/EL, May 1994.
- [6] C. Elachi, *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE Press, New York, 1998.
- [7] T. M. Lillesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*. John Wiley & Sons, 2nd Edition, 1987.
- [8] P. M. Mather, *Computer Processing of Remotely-Sensed Images*, John Wiley & Sons, 1987.
- [9] C. Oliver and S. Quegan S., *Understanding Synthetic Aperture Radar Images*, Artech-House, London, 1998.
- [10] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*. Springer-Verlag, Berlin, 1999.
- [11] F. F. Sabins, *Remote sensing: Principles and interpretation*. W. H. Freeman and Company, New York, 2nd Edition, pp. 235-277, 1987.
- [12] L. Bruzzone and S. B. Serpico, "Detection of changes in remotely-sensed images by the selective use of multi-spectral information", *International Journal of Remote Sensing*, Vol. 18, No. 18, pp. 3883-3888, 1997.
- [13] M. J. Carlloto, "Detection and analysis of change in remotely-sensed imagery with application to wide area surveillance", *IEEE Transaction on Image Processing*, Vol. 6, No. 1, pp. 189-202, 1997.
- [14] A. T. Chrien, R. Green, and M. Eastwood, "Accuracy of the spectral and radiometric laboratory calibration of the airborne visible infrared imaging spectrometer (AVIRIS)", In *Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop*, June 1990.
- [15] C. Lee and D. Landgrebe, *Feature Extraction and Classification Algorithms for High Dimensional Data*, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, TR-EE 93-1, January 1993.
- [16] J. P. Hoffbeck and D. Landgrebe, *Classification of High Dimensional Multispectral Data*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-EE95-14, May 1995.
- [17] Q. Z. Jackson and D. Landgrebe, *Design of an Adaptive Classification Procedure for the Analysis of High-Dimensional Data with Limited Training Samples*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-ECE 01-5, December 2001.

- [18]. W. M. Boerner *et al.*, "Polarimetry in radar remote sensing: Basic and applied concepts," in *The Manual of Remote Sensing*, 3rd Edition, New York: Wiley & Sons, 1998, Ch. 5.
- [19] E. R. Stofan, D. L. Evans, C. Schmullius, B. Holt, J. J. Plaut, J. Van Zyl, S. D. Wall, J. Way, "Overview of results of spaceborne imaging radar-C, X-band synthetic aperture radar", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 33, No. 4, pp. 817-828, 1995.
- [20] J. Rolando, L. Bryan, L. Huneycutt, W. Marian, "The SIR-C/X SAR synthetic aperture radar system", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 33, No. 4, pp. 829-839, 1995.
- [21] N. H. Kenea, "Improved geological mapping using Landsat TM data, Southern Red Sea Hills, Sudan: PC and IHS decorrelation stretching", *International Journal of Remote Sensing*, Vol. 18, No. 6, pp. 165-177, 1997.
- [22] Jr. P. S. Chavez and J. A. Bowell, "Comparison of the spectral information content of Landsat Thematic Mapper and SPOT for three different sites in the Phoenix, Arizona region", *Photogrammetric Engineering and Remote Sensing*, Vol. 54, No. 12, pp. 1699-1708, 1988.
- [23] B. Guarguet-Duport, J. Girel, J. M. Chassery, and G. Pautou, "The use of multiresolution analysis and wavelets transform for merging SPOT panchromatic and multispectral image data", *Photogrammetric Engineering and Remote Sensing*, Vol. 62, No. 9, pp. 1057-1066, 1996.
- [24] I. Lakkis, S. El-Assad, and J. Saillard, "Traitement optimaux pour l'imagerie radar à ouverture synthétique", *Revue Traitement du Signal*, Vol. 11, No. 1, pp. 3-14, 1994.
- [25] Ch. Barbier. "Systèmes spatiaux de télédétection", *Support de cours*, Centre Spatial de Liège, Liège, Belgique, 1997.
- [26] Y. Chibani, *Méthodes et Algorithmes de Fusion d'Images Satellitaires Dédiées à l'Amélioration de l'Interprétation*. Thèse de Doctorat d'Etat, Faculté d'Electronique et d'Informatique, Université des Sciences et Technologie Houari Boumediene U. S. T. H. B., Alger Octobre 2001.
- [27] G. Celeux, E. Diday, G. Covaert, et Y. Lechevalier, *Classification Automatique des Données*. Dunod Informatique, 1989.
- [28] G. Saporta, *Probabilités Analyse de Données et Statistiques*. Editions Technip, 1990.
- [29] W. Pieczynski, "Estimation of context in random fields", *Journal of Applied Statistics*, Vol. 19, No. 2, pp. 283-289, 1989.
- [30] O. Pony, X. Descombes, J. Zerubia, "Classification d'images satellitaires hyperspectrales en zones rural et périurbaine", *Rapport de recherche No. 4008*, Institut National de Recherche en Informatique et en Automatique, Sophia Antipolis, France., 2000.
- [31] A. H. Schistad Solberg, T. Taxt, A. K. Jain, "A Markov random field model for classification of multisource satellite imagery", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 34, No. 1, pp. 100-113, 1996.
- [32] A. H. Schistad Solberg, "Contextuel data fusion applied to Forest map revision", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, No. 3, pp. 1234-1243, 1999.
- [33] L. Wald, "Some terms of reference in data fusion", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, No. 3, pp. 1190-1193, 1999.
- [34] L. Wald, "A conceptual approach to the fusion of Earth observation data", *Survey in Geophysics*, Vol. 21, pp. 177-186, 2000.
- [35] N. Khazenie, M. M. Crawford, "Spatial-temporel autocorrelated model for contextuel classification", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 28, No. 4, pp. 529-539, 1990.
- [36] J. Byeungwoo, D. A. Landgrebe, "Classification with spatio-temporel interpixel class dependency contexts", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 30, No. 4, pp. 663-672, 1992.
- [37] J. Byeungwoo, D. A. Landgrebe, "Decision fusion approach for multitemporel classification", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, No. 3, pp. 1227-1233, 1999.
- [38] W. Pratt, *Digital Image processing*. John Willey and sons, New York, 750 p., 1978.
- [39] A. Belhadj-Aissa, *Contribution au Logiciel d'Analyse et de Traitement d'Images Satellitaires (L.A.T.I.S). Analyse Spatial-Spectral d'Images Satellitaires Appliquées à la Cartographie*

- Thématique*. Thèse de Doctorat d'Etat en Electronique des Systèmes, spécialité Traitement d'Images et Télédétection, Université des Sciences et Technologie Houri Boumediène, Alger, 268 p., 1998.
- [40] S. Chitroub, R. Khedam, H. Belhadj, et B. Sansal, "Clasification contextuelle d'images utilisant les Matrices de Co-occurences de Niveaux de Gris (MCNG) ", In *Proceedings of Conférence Maghrébine sur l'Automatique, l'Electrotechnique et l'Electronique Industriel (COMAEI)*, Béjaïa, du 7 au 8 Decembre, 1998, Algérie, pp. 105-108.
 - [41] H. Kaizer, "A Quantification of textures on aerial photographs", *Technical Note 121, AD 69484*, Boston University, 1955.
 - [42] M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification", *IEEE Transactions on Systems, Man, Cybernetics*, Vol. SMC-3, No. 6, pp. 610-621, 1973.
 - [43] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. L. Williams, "Textural informartion in SAR images", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 24, No. 3, pp. 235-245, 1986.
 - [44] M. Haralick and L. G. Shapiro, "Glossary of computer vision terms", *Pattern Recognition*, Vol. 24, No. 5, pp. 391-399, 1991.
 - [45] A. Baraladi and F. Parmiggiani, "An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 33, No. 2, pp. 293-304, 1995.
 - [46] L. Wang and D. C. He, "A statitical approach for texture analysis", *Photogrammetric Engineering and Remote Sensing*, Vol. 56, No. 1, pp. 61-66, 1990.
 - [47] D. C. He and L. Wang, "Texture features based on texture spectrum", *Pattern Recognition*, Vol. 24, No. 5, pp. 69-93, 1991.
 - [48] D. C. He and L. Wang, "Unsupervised textural classification of images using the texture spectrum", *Pattern Recognition*, Vol. 25, No. 3, pp. 274-255, 1992.
 - [49] R. Chevalier, A. Fontanel, g. Grau, and M. Guy, "Application du filtrage optique à l'étude des photographies aeriennes", *11th Congress International Photogramm*, 1968.
 - [50] J. Huang and D. L. Turcotte, "Fractal mapping of digitized images: application to the topography of Arizon and comparison with synthetic images", *Journal of Geophysics Reasch*, Vol. 94, No. B6, pp. 7491-7495, 1989.
 - [51] D. L. Turcotte, *Fractals and chaos in geology and geophysics*. Cambridge University Press, 221p, 1992.
 - [52] M. P. Dubuisson and R. C. Dubes, "Efficacy of fractal features in segmenting images of natural textures", ", *Pattern RecognitioLetters*, Vol. 15, pp. 419-431, 1994.
 - [53] G. R. Cross and A. K. Jain, "Markov random field texture models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-5, No. 1, pp. 25-39, 1983.
 - [54] R. Hu and M. M. Fahmy, "Texture segmentation based on a hierarchical Markov random field model", *Signal Processing*, Vol. 26, pp. 285-305, 1992.
 - [55] J. Mao and A. K. Jain, "Texture classification and segmentation using multiresolution simultaneous autoregressive models", *Pattern Recognition*, Vol. 25, No. 2, pp. 173-188, 1992.
 - [56] T. Ranchin, B. Naert, M. Albuisson, G. Boyer, and P. Astrand, "An Automatic method for vine detection in airborne imagery using wavelet transform and multiresolution analysis", *Photogrammetric Engineering and Remote Sensing*, Vol. 67, No. 1, pp. 91-98, 2001.
 - [57] Y. Meyer, *Les Ondelettes : Algorithmes et Applications*. Armand Colin Editeurs, Paris, France, 171p., 1992.
 - [58] M. Mangolini, *Apport de la fusion d'images satellitaires multicapteurs au niveau pixels en télédétection et photo-interprétation*. Thèse de Doctorat en Sciences, Université de Nice Sophia Antipolis, Nice, France, 174 p., 1994.
 - [59] T. Ranchin, "Wavelets for modeling in remote sensing", In *Proceedings of the SPIE, Conference on wavelet applications in Signal and image Processing VII*, Denver, Colorado, Vol. 3813, 1999, pp. 211-214.
 - [60] C. Zhu and X. Yang, "Study of remote sensing image texture analysis and classification using waveltes", *International Journal of Remote Sensing*, Vol. 19, No. 16, pp. 3197-3203, 1998.

- [61] C. J. S. Ferro, *Scale and texture in digital classification*. Thesis MSC to the Eberly College of Arts and Sciences at West Virginia University in partial fulfillment of the degree of Master of Arts in Geography, Morgantown, West Virginia, 52 p., 1998.
- [62] J. Desachy, *Interpretation automatique d'images satellitaires: le système ICARE*. Thèse de Doctorat en Informatique, Université Paul Sabatier, Toulouse, France, 1991.
- [63] C. Pohl and J. L. Van Genderen, "Multisensor fusion in remote sensing: Concepts, methods and applications", *International Journal of Remote Sensing*, Vol. 19, No. 5, pp. 823-854, 1998.
- [64] T. Ranchin, *Application de la transformée en ondelettes et de l'analyse multi-résolution au traitement des images de télédétection*. Thèse de Doctorat en Sciences, Université de Nice Sophia Antipolis, Nice, France, 110 p., 1993.
- [65] C. Pohl, *Geometric aspects of multisensor image fusion for topographic map updating in the humid tropics*. Doktor Thesis, Universität Hannover, 159p, 1996.
- [66] S. Chauvin, *Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire*. Thèse de Doctorat, Université de Nantes, Nantes, France, 222 p., 1995.
- [67] P. Burt and A. E. Adelson, "The Laplacian pyramid as a compact image code", *IEEE Transactions on Communications*, Vol. 31, pp. 532-540, 1983.
- [68] A. Toet, "Image fusion by a ratio of low-pass pyramid", *Pattern Recognition Letters*, Vol. 9, pp. 245-253, 1989.
- [69] A. Toet, "Multiscale contrast enhancement with applications to image fusion", *Optical Engineering*, Vol. 31, pp. 1026-1031, 1992.
- [70] A. Toet, "Multiscale color image enhancement", *Pattern Recognition Letters*, Vol. 13, pp. 167-174, 1992.
- [71] S. G. Mallat, "A multiresolution signal decomposition: The wavelet transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693, 1989.
- [72] L. J. Chipman and T. M. Orr, "Wavelets and fusion", In *Proceedings of International Congress on Image Processing*, Vol. 1, pp. 248-251, 1995.
- [73] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform", *Graphical Models and Image Processing*, Vol. 57, No. 3, pp. 235-245, 1995.
- [74] T. A. Wilson, S. K. Rogers, and L. R. Myers, "Perceptual-based hyperspectral image fusion using multiresolution analysis", *Optical Engineering*, Vol. 34, pp. 3145-3164, 1995.
- [75] T. Ranchin and L. Wald, "Comparison of different algorithms for the improvement of the spatial resolution of images", In *Proceedings of The Third International Conference on Fusion of Earth Data: "Merging point measurements raster maps and remotely sensed images"*, Sophia Antipolis, France, 26-28 January, 2000, pp. 97-98.
- [76] T. Ranchin and L. Wald, "Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation", *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 1, pp. 49-61, 2000.
- [77] J. Nùñez, X. Otazu, O. Fors, A. Prades, V. Palà, R. Arbiol, "Multiresolution-based image fusion with additive wavelet decomposition", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, No. 3, pp. 1204-1211, 1999.
- [78] C. Lartry, "Restauration des images optiques satellitaires", *Séminaire Restauration d'Images de Télédétection, CCT TSI, CNES, Toulouse, France, Avril 2001*.
- [79] A. K. Jain, *Fundamentals of Digital Image Processing*. Information and Systems Sciences series, Prentice Hall, 1989.
- [80] A. Jalobeanu, *Modèles, estimation Bayésienne et algorithmes pour la déconvolution d'images satellitaires et aériennes*. Thèse de Doctorat, Spécialité Traitement des images, Ecole Doctorale de Sciences Fondamentales et Appliquées, Université de Nice-Sophia Antipolis, Nice, France, décembre 2001.
- [81] W. M. Brown and L. J. Porcello, "An introduction to synthetic aperture radar", *IEEE Spectrum*, pp. 52-62, 1969.
- [82] I. Lakkis, S. El-Assad, et J. Saillard, "Traitements optimaux pour l'imagerie radar à ouverture synthétique", *Traitement du Signal*, Vol. 11, No. 1, pp. 3-14, 1994.
- [83] K. Tomiyasu, "Tutorial review of synthetic aperture radar (SAR) with applications to imaging of the ocean surface", In *Proceedings of the IEEE*, Vol. 66, No. 5, pp. 563-583, 1978.

-
- [84] F. T. Ulaby, "Radar signatures of terrain : Useful monitors of renewable resources", In *Proceedings of the IEEE*, Vol. 70, No. 12, pp. 1410-1428, 1982.
- [85] J. W. Goodman, *Laser speckle and related phenomena*. Edition Dainty, J. C., Springer, New York, pp. 10-75, 1975.
- [86] Ph. Olivier, "Principes de l'imagerie haute resolution au moyen d'un radar à ouverture synthétique embarqué sur satellite", *Note Technique CRPE/ 179*, décembre 1989.
- [87] S. Chitroub, A. Houacine, and B. Sansal, "Supervised fusion-classification of multispectral images using fuzzy sets theory", *IEE International Conference on Image Processing and its Application (IPA'99)*, 13-17 July 1999, Manchester, England, pp. 96-97.
- [88] S. Chitroub A. Houacine et B. Sansal, "Fusion et classification d'image multispectrale en utilisant la logique floue", *DAT 2000*, du 21 au 23 mai, 2000, Reghai, Algérie pp.57-63.
- [89] S. Chitroub, A. Houacine, and B. Sansal, "Evidential reasoning-based classification method for remotely sensed images", In *Proceedings of SPIE, Image and Signal Processing for Remote Sensing Conference*, 18-21 September 2001, Toulouse, France, Vol. 4541, pp. 340-351.
- [90] H. H. Arsenault and G. V. April, "Information content of images degraded by speckle noise", *Optical Engineering*, Vol. 25, No. 5, pp. 662-666, 1989.
- [91] J. W. Goodman, "Some fundamental properties of speckle", *Journal of Optical Society of America*, Vol. 66, No. 11, pp. 1145-1149, 1976.
- [92] F. Kimura and M. Shridhar, "Handwritten numerical recognition based on multiple algorithms", *Pattern Recognition*, Vol. 24, No. 10, pp. 969-983, 1991.
- [93] B. Moobed, *Combinaison de classifieurs, une nouvelle approche*. Thèse présentée pour obtenir le grade de Docteur en science, Spécialité Informatique, Ecole Polytechnique LIX, Université de Paris Sud UFR Scientifique d'Orsay, France, 1996.
- [94] A. Serir, *Analyses multirésolution et multiéchelle appliquées aux traitements d'images satellitaires*. Thèse de Doctorat d'Etat, Faculté d'Electronique et d'Informatique, Université des Sciences et Technologie Houari Boumediene U. S. T. H. B., Alger 2002.

Chapitre 2

Caractérisation Statistique et Modélisation d'Images Optiques et Radars

1. Introduction

La caractérisation statistique et la modélisation d'images de télédétection sont d'une importance capitale dans le développement des algorithmes d'analyse et de classification. Nous présentons ici les modèles probabilistes qui décrivent les comportements statistiques des images fournies par les satellites de télédétection civile, optiques et radars. L'exactitude des estimateurs des paramètres des modèles est un problème crucial qui dépend de plusieurs facteurs en particulier de l'opération d'extraction des échantillons d'apprentissage. Nous exposons les différentes méthodes d'estimation des paramètres statistiques des modèles ainsi que les manières d'obtenir des estimateurs précis.

Pour les images multispectrales optiques, il a été prouvé que la distribution Gaussienne décrit d'une manière fiable le comportement spectral des réponses réfléchies par les objets terrestres [1]-[3]. Cette distribution est complètement caractérisée par les statistiques d'ordre un et deux. Le problème majeur dans l'utilisation de ce modèle est l'estimation statistique de ces paramètres. Sans tenir compte du type d'estimateur utilisé, l'exactitude de l'estimation est fortement dépendante d'une part de la représentativité des échantillons d'apprentissage et d'autre part du nombre de ces échantillons. En pratique, l'extraction d'un nombre suffisant d'échantillons fiables est une tâche fastidieuse [4]. Le paragraphe deux de ce chapitre est consacré à la modélisation des images multispectrales optiques par la loi Gaussienne. Les critères de sélection des échantillons d'apprentissage sont donnés ainsi que l'estimation des statistiques d'ordre un et deux. Le cas d'images multispectrales de dimension élevée est aussi abordée en mettant l'accent sur le problème de l'estimation de la statistique d'ordre deux dans les conditions d'insuffisance de nombre d'échantillons d'apprentissage. De nouveaux estimateurs de covariance, développés récemment dans la littérature sont alors présentés.

En ce qui concerne les images radars à ouverture synthétique (SAR), plusieurs modèles probabilistes ont été proposés dans la littérature. Au début de l'exploitation des images SAR, les distributions telles que Lognormale, Weibull, Beta modifiée, et Gamma ont été testées [5]-[8]. Des chercheurs ont néanmoins découvert que ces distributions ne décrivent pas convenablement les comportements des signaux radars réfléchis [9]-[11]. Récemment, la distribution K et la distribution complexe de Wishart ont été proposées. La distribution K fait l'objet de plusieurs publications. Cependant, on ne trouve pas beaucoup de chercheurs qui travaillent sur l'utilisation de la distribution de Wishart comme modèle pour les images SAR. L'utilisation restrictive de ce modèle peut être due d'une part, à l'absence d'un fondement théorique convaincant sur la convenance du modèle pour les images SAR et d'autre part, aux difficultés pratiques rencontrées dans la mise en œuvre du modèle en raison du nombre et la nature des paramètres à estimer [12]-[16]. En revanche, la distribution K a été utilisée intensivement par les chercheurs [9]-[11], [17]-[22]. Elle présente l'avantage d'être théoriquement justifiée comme un modèle décrivant fidèlement les réponses des signaux radars réfléchis. Sa mise en œuvre se réduit à estimer un seul paramètre à la fois en utilisant des méthodes d'estimations simples à implémenter [11]. Le paragraphe 3 apporte plus de détails sur la modélisation des images SAR par la distribution de Wishart et la distribution K. L'utilisation restrictive de la distribution de Wishart ainsi que la popularité de la distribution K sont justifiées. L'estimation des paramètres de ces deux distributions est aussi abordée.

2. Caractérisation Statistique et Modélisation d'Images Optiques

En télédétection spatiale, les ondes électromagnétiques émises, réfléchies ou diffractées par les objets terrestres ou l'atmosphère sont utilisées comme support d'information sur l'état de la surface observée par le satellite. Les capteurs optiques permettent d'acquérir des réponses liées à la signature spectrale électromagnétique de l'objet observé suite à un rayonnement solaire. Cette signature dépend essentiellement de la nature et de l'état de l'objet au moment de l'observation ainsi que des conditions de propagation des ondes électromagnétiques réfléchies par celui-ci et son environnement [3], [4]. Dans ce qui suit, nous allons parler de la caractérisation statistique de l'image multispectrale optique via la modélisation probabiliste de cette dernière.

2. 1. Modélisation par la Distribution Gaussienne

Chaque point élémentaire d'une image multispectrale optique possède comme signature spectrale un vecteur x de dimension k , où k représente le nombre de bandes spectrales du système d'acquisition embarqué sur le satellite. Les composantes du vecteur x sont les valeurs radiométriques du pixel dans chaque bande spectrale. Ces valeurs radiométriques

représentent les intensités des rayonnements captés par les capteurs. Il a été prouvé que le comportement spectral d'un objet terrestre (classe de la couverture terrestre) est bien décrit par la distribution Gaussienne multidimensionnelle [1]-[4], telle que :

$$P(x|w_i) = \frac{1}{\sqrt{(2\pi)^k |S_i|}} \cdot \exp\left(-\frac{1}{2} (x - m_i)^T S_i^{-1} (x - m_i)\right) \quad (1)$$

où m_i et S_i sont est le vecteur moyenne et la matrice de covariance, respectivement, de la classe w_i . En pratique, les paramètres m_i et S_i sont inconnus et sont estimés à partir des échantillons d'apprentissage.

2. 1. 1. Sélection des échantillons d'apprentissage

La sélection des échantillons d'apprentissage nécessite l'utilisation des données de référence et une connaissance *a priori* de la scène. Les données de référence peuvent être les cartes thématiques, les images classifiées ou des mesures réalisées sur le terrain. Ces mesures sont désignées par les vérités terrain. Le nombre et la représentativité des échantillons sont deux facteurs qui influent sur l'exactitude des estimateurs des paramètres statistiques du modèle Gaussien. La taille de l'ensemble des échantillons est liée au nombre de bandes spectrales et au nombre de paramètres à estimer [23]. Théoriquement, la taille minimale d'un ensemble d'échantillons est $k+1$ par classe. Une taille inférieure peut engendrer une matrice de covariance singulière de la classe correspondante. Pratiquement, si nous avons une bande spectrale et une statistique à estimer, un ensemble de 10 à 100 échantillons par classe est considéré comme suffisant pour assurer une bonne estimation [23]. Pour un nombre de k bandes spectrales, un ensemble minimum de $10 \times k$ échantillons par classe est nécessaire [23], [24]. Dans tous les cas, il est préférable d'avoir autant d'échantillons d'apprentissage pour chaque classe afin d'assurer des estimateurs exacts de ses paramètres statistiques.

Le deuxième facteur influant sur l'exactitude des estimateurs est la qualité des échantillons sélectionnés (degré de la représentativité de ces échantillons aux variations spectrales des classes). Le critère de dispersion de l'extraction des échantillons à travers toute la scène est alors nécessaire. Ce critère est justifié par deux raisons : La première est l'effet d'autocorrélation spatiale des pixels adjacents. En effet, la taille de l'ensemble des échantillons spécifiée est valable sous l'hypothèse que les échantillons sélectionnés soient indépendants. Les pixels adjacents ne sont pas indépendants, ils présentent une autocorrélation spatiale [23]. Si les échantillons sélectionnés présentent une autocorrélation spatiale plus au moins importante, il s'ensuit qu'on a une surestimation de l'information réelle caractérisant le comportement spectral de la classe. Cela est dû au fait que les pixels adjacents ne décrivent pas réellement la variation spectrale à l'intérieur de la classe. Il a été

prouvé [23] que l'estimation de la matrice de covariance à partir des échantillons sélectionnés aléatoirement dans la classe est plus précise que celle estimée à partir des échantillons extraits des blocs contigus des pixels.

La deuxième raison consiste à essayer d'assurer au maximum la représentativité des échantillons aux variations spectrales de la classe. En effet, une classe peut se manifester, en même temps, en plusieurs formes spectrales proches et dans des endroits différents de l'image. Pour cela, nous avons adopté deux définitions différentes pour une classe, classe d'information et classe spectrale. Une classe d'information est constituée de plusieurs classes spectrales. Une classe spectrale est un comportement spectral différent d'un même type terrestre. Pour mieux caractériser le comportement spectral de l'objet terrestre par des paramètres statistiques, nous devons estimer ces paramètres à partir des échantillons tirés de différentes classes spectrales constituant la classe d'information de l'objet terrestre.

2. 1. 2. Estimation du vecteur moyenne et de la matrice de covariance

Deux types d'estimation des paramètres peuvent être utilisés, estimation paramétrique et estimation non paramétrique. Si on suppose que la fonction de densité de probabilité, "*Probability Density Function (PDF)*", de la classe a une forme connue et que seuls, les paramètres de cette densité doivent être estimés, l'estimation est dite paramétrique. Si ce n'est pas le cas, on dit que l'estimation est non paramétrique. Dans le cas où les classes sont connues *a priori*, l'estimation (paramétrique ou non) est dite supervisée. Dans le cas contraire, l'estimation est dite non-supervisée.

Dans le cadre de l'estimation paramétrique supervisée, on cherche donc à résoudre le problème suivant : Utilisant les échantillons tirés de la classe w_i , on estime la PDF de cette classe en supposant que sa forme est connue et dépend des paramètres statistiques a_i . Pour les images multispectrales optiques, la forme de la PDF est Gaussienne. Donc les paramètres a_i à estimer sont le vecteur moyenne et la matrice de covariance de la classe. Plusieurs méthodes d'estimation peuvent être utilisées telles que la méthode d'estimation au sens de maximum de vraisemblances, méthode Bayésienne, méthode en utilisant la notion de statistique suffisante, et méthode d'analyse discriminante régularisée [2], [23]-[25]. La méthode la plus utilisée est la méthode d'estimation au sens de maximum de vraisemblances qui consiste à maximiser par rapport aux paramètres a_i la probabilité d'obtenir les échantillons x sachant a_i . Cette probabilité est considérée comme une fonction des paramètres a_i , notée $P(x|a_i)$. Admettant que les N_i échantillons tirés de la classe w_i sont indépendants les uns des autres, nous avons alors :

$$P(x|a_i) = \prod_{k=1}^{N_i} P(x_k|a_i) \quad (2)$$

Maximiser $P(x|a_i)$ par rapport à a_i revient aussi à maximiser son logarithme $L(a_i)$ (puisque cette dernière est une fonction croissante monotone) :

$$L(a_i) = \log(P(x|a_i)) = \sum_{k=1}^{N_i} \log(P(x_k|a_i)) \quad (3)$$

Une condition nécessaire pour que $L(a_i)$ soit maximale est que :

$$\nabla_{a_i}(L(a_i)) = \sum_{k=1}^{N_i} \nabla_{a_i}[\log(P(x_k|a_i))] = 0 \quad (4)$$

où l'opérateur ∇_{a_i} représente le gradient. Pour le cas de la PDF Gaussienne monovariante, maximiser la vraisemblance d'un échantillon x_k revient alors à maximiser l'expression :

$$\log(P(x_k|m_i, v_i)) = -\frac{1}{2} \cdot \log(v_i) - \frac{1}{2 \cdot v_i} (x_k - m_i)^2 \quad (5)$$

où m_i et v_i sont la moyenne et la variance de la PDF de la classe w_i .

Soient $a_1 = m_i$ et $a_2 = v_i$, la condition d'extremum $\nabla_{a_i}(L(a_i)) = 0$ prend la forme :

$$\sum_{k=1}^{N_i} \begin{pmatrix} \frac{1}{a_2} (x_k - a_1) \\ -\frac{1}{2 \cdot a_2} + \frac{(x_k - a_1)^2}{2 \cdot a_2^2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad (6)$$

Les estimateurs $\hat{a}_1$ et $\hat{a}_2$ des paramètres a_1 et a_2 sont les solutions du système d'équations :

$$\begin{cases} \sum_{k=1}^{N_i} \frac{1}{\hat{a}_2} (x_k - \hat{a}_1) = 0 \\ -\sum_{k=1}^{N_i} \frac{1}{2 \cdot \hat{a}_2} + \sum_{k=1}^{N_i} \frac{(x_k - \hat{a}_1)^2}{2 \cdot \hat{a}_2^2} = 0 \end{cases} \quad (7)$$

En notant $m_i = \hat{a}_1$ et $v_i = \hat{a}_2$, nous obtenons :

$$\hat{m}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} x_k \quad (8)$$

$$\hat{v}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (x_k - \hat{m}_i)^2 \quad (9)$$

Dans le cas Gaussien multivariante, l'estimation du vecteur moyenne, m_i , et de la matrice de covariance, S_i , de la classe w_i est en tout point semblable à l'estimation dans le cas Gaussien monovariante. Elle nécessite cependant des calculs beaucoup plus lourds et ne sera pas présentée ici. Nous arrivons aux résultats bien connus:

$$\hat{m}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} x_k \quad (10)$$

$$\hat{S}_i = \frac{1}{N_i} \sum_{k=1}^{N_i} (x_k - \hat{m}_i)(x_k - \hat{m}_i)^T \quad (11)$$

Bien que biaisés, les estimateurs donnés dans (8), (9), (10), et (11) sont très couramment utilisés car ils sont simples dans la mise en œuvre.

2. 2. Cas d'Images Multispectrales de Dimension Elevée

La nouvelle technologie des systèmes d'acquisition de télédétection spatiale a permis de fournir des images multispectrales correspondant à des centaines de bandes spectrales. Le système aéroporté AVIRIS, par exemple, fournit des images dans 220 bandes spectrales [26]-[31]. Cette amélioration dans la résolution spectrale devrait fournir une richesse d'informations détaillées. Malheureusement, les techniques utilisées pour l'analyse d'images multispectrales de dimension faible ne sont pas valables pour analyser les images multispectrales de dimension élevée [2], [26]-[31]. Par exemple, pour la modélisation d'images, le nombre d'échantillons d'apprentissage nécessaire à l'estimation des paramètres du modèle Gaussien multidimensionnel de dimension élevée est beaucoup plus grand que dans le cas de dimension modérée. L'extraction de ces échantillons devient très difficile et coûteuse. Par conséquent, le critère d'avoir assez d'échantillons d'apprentissage pour assurer une estimation précise des statistiques du modèle, risque d'être impossible dans le cas d'images multispectrales de dimension élevée. Autrement dit, afin d'exploiter la richesse d'informations contenue dans ce type d'images, de nouveaux algorithmes d'analyse doivent être développés. Ainsi, proposer de nouvelles méthodes d'estimation des paramètres du modèle Gaussien de dimension élevée devient une nécessité primordiale, en particulier pour l'estimation de la matrice de covariance.

2. 2. 1. Estimation du vecteur moyenne

Il a été prouvé dans plusieurs études que, malgré la dimension élevée de l'image multispectrale, l'estimation du vecteur moyenne n'est pas très sensible au nombre d'échantillons d'apprentissage [2], [4], [26], [28], [31]. L'utilisation de l'estimateur au sens de maximum de vraisemblances, donné dans l'expression (10), est très recommandée dans la littérature en raison de sa simplicité de mise en œuvre. Cependant, afin de rendre l'estimateur non biaisé, la notion de vecteur moyenne échantillonné est utilisée et l'expression (10) devient alors :

$$m_i = \frac{1}{N_i - 1} \sum_{k=1}^{N_i} x_k \quad (12)$$

2. 2. 2. Estimation de la matrice de covariance

L'estimation de la matrice de covariance dans le cas d'images multispectrales de dimension élevée fut l'objet de plusieurs publications et thèses. Ceci est dû, d'une part, à la sensibilité de l'estimateur de covariance au nombre et qualité des échantillons d'apprentissage, et d'autre part, à l'importance de la matrice de covariance dans le développement et la performance de plusieurs méthodes d'analyse et de classification d'images optiques multidimensionnelles. Si le nombre d'échantillons d'apprentissage est petit, la covariance estimée est fortement elliptique et l'écart par rapport à la covariance réelle est important [32]. En revanche, la covariance commune estimée peut mener à une exactitude plus élevée, même lorsque les vraies matrices de covariance des classes sont tout à fait différentes [25]. La robustesse d'un tel estimateur est alors dépendante des vraies statistiques des classes, de la dimensionalité et du nombre d'échantillons d'apprentissage. Il est utile alors de déterminer si la covariance échantillonnée, commune, ou toute autre type de covariance estimée est appropriée dans une telle situation. Ceci est le but des nouveaux estimateurs présentés dans ce sous paragraphe. Avant d'aborder ces nouveaux estimateurs, nous allons donner d'abord l'état de l'art des différents estimateurs proposés dans la littérature.

2. 2. 2. a- Etat de l'art :

Dans la pratique, la matrice de covariance est souvent estimée par la matrice de covariance échantillonnée dont l'expression est dérivée de l'estimateur au sens de maximum de vraisemblances donné dans l'expression (11) telle que [24]:

$$S_i = \frac{1}{N_i - 1} \sum_{k=1}^{N_i} (x_k - \hat{m}_i)(x_k - \hat{m}_i)^T \quad (13)$$

Cependant, lorsque la dimensionalité des données est grande, la variance de l'estimateur (13) est élevée et la matrice peut être singulière si le nombre d'échantillons d'apprentissage est inférieur à la dimensionalité. Une manière de palier à ce problème est de réduire le nombre de paramètres à estimer en remplaçant les matrices de covariance échantillonnée des classes par leur moyenne, S_w , telle que :

$$S_w = \frac{1}{N - M} \sum_{i=1}^M (N_i - 1) S_i \quad (14)$$

où N est le nombre total d'échantillons d'apprentissage et M le nombre de classes. Ainsi, en utilisant la matrice moyenne S_w , le nombre de paramètres est sensiblement réduit, et les variances des éléments de S_w sont plus petites que celles des éléments correspondants des matrices S_i . Quoique chaque matrice de covariance S_i puisse différer sensiblement, les

études menées dans [4], [26], [27], [33], [34] et [35] ont prouvé que la diminution des variances des estimateurs des paramètres mène souvent à améliorer les performances de la classification dans le cas d'un nombre d'échantillons limité. Plusieurs méthodes de mise en commun des matrices de covariance de formes diverses ont été proposées [25], [35], [26], [31]. Le principe est d'optimiser, via des paramètres de régularisation, l'opération de la mise en commun des matrices de covariance en réduisant autant que possible le nombre de paramètres à estimer. L'optimisation s'effectue en deux phases : Phase destinée à choisir les modèles de mise au commun des matrices de covariance et phase pour sélectionner un modèle approprié selon les paramètres de régularisation.

Pour la première phase, on doit donc décider sur un ensemble de modèles appropriés pour la mise au commun des covariances qui représente un ensemble plausible de matrices de covariances estimées. La forme de mise en commun des covariances est donnée par :

$$\hat{S}_i = (1 - \alpha_i)S_i + \alpha_i C \quad 0 \leq \alpha_i \leq 1 \quad (15)$$

L'optimisation via le paramètre de mise au commun, α_i , permet donc de contrôler le biais de chaque estimateur de matrice de covariance échantillonnée S_i par rapport à la matrice de covariance commune C . Cependant, cette mise en commun partielle des covariances peut ne pas fournir assez de régularisation [31], en particulier lorsque le nombre total d'échantillons N est comparable ou inférieur à la dimension de l'image multispectrale p . Dans ces conditions, la matrice de covariance commune C devient problématique. Une variante de mise en commun des covariances est fournie en biaisant les covariances échantillonnées avec une certaine matrice diagonale non singulière L :

$$\hat{S}_i = (1 - \alpha_i)S_i + \alpha_i L \quad 0 \leq \alpha_i \leq 1 \quad (16)$$

Pour des valeurs données du paramètre α_i , la présence du terme biaisé dans les expressions (15) et (16) dépendra du degré de représentativité de la matrice estimée $\hat{S}_i$ à la vraie matrice S_i .

La deuxième phase consiste en la sélection d'un modèle de mise en commun des covariances. Cette sélection consiste à faire un choix optimal de la valeur du paramètre α_i . Ceci est possible en considérant l'estimation de ce paramètre comme étant une fonction de perte à minimiser en utilisant les échantillons d'apprentissage. Un critère populaire de minimisation est basé sur l'évaluation de l'erreur de classification. L'idée est d'utiliser le classifieur conçu en utilisant les $N_i - 1$ échantillons en excluant l'échantillon $x_{i,k}$ pour classifier $x_{i,k}$, un paramètre est alors choisi pour réduire au minimum le taux d'erreur de classification. Bien que l'utilisation de ce critère ait l'avantage d'être directement lié à l'exactitude de classification, elle présente cependant quelques inconvénients. Tout d'abord, du point de vue temps de calcul, elle est coûteuse. En second lieu, le même

paramètre de mélange doit être utilisé pour toutes les classes puisque le procédé de classification utilise simultanément toutes les matrices de covariance des classes. Cependant, le choix unique du paramètre de mélange ne pourrait pas être optimal pour toutes les classes. En outre, le même taux d'erreur de classification pourrait se produire le long d'un éventail de valeurs du paramètre et par conséquent la valeur optimale du paramètre de mélange n'est pas unique. Un autre critère, qui sera détaillé ultérieurement, est de maximiser la somme des valeurs de la moyenne des probabilités de vraisemblance des échantillons non utilisés dans le processus de régularisation. Ce critère exige moins de calcul que le critère lié à l'erreur de classification. Il a également l'avantage que chaque matrice de covariance de classe peut être estimée indépendamment des autres. Par conséquent, le paramètre de mélange peut être différent pour chaque classe. De plus, ce ne sont pas toutes les classes qui sont soumises à la régularisation, particulièrement celles ayant un nombre d'échantillons d'apprentissage suffisant. Cependant, un inconvénient majeur de ce critère est le manque de rapport direct avec l'exactitude de classification.

Selon le principe de mise en commun optimal des covariances, plusieurs méthodes ont été proposées dans la littérature. Parmi ces méthodes, il y a la méthode d'analyse discriminante régularisée, "*Regularized Discriminant Analysis (RDA)*", qui optimise la mise en commun des covariances selon les expressions suivantes [25] :

$$\hat{S}_i(l, g) = (1 - g) \cdot \hat{S}_i(l) + g \left(\frac{\text{tr}(\hat{S}_i(l))}{p} \right) I \quad 0 \leq g \leq 1 \quad (17)$$

$$\hat{S}_i(l) = \frac{(1 - l)(N_i - 1)S_i + l(N - M)S_w}{(1 - l)N_i + lN} \quad 0 \leq l \leq 1 \quad (18)$$

I est la matrice identité. L'aspect régularisation de la méthode est de maximiser un indice d'optimisation, qui est lié directement à l'erreur de classification, via les paramètres de mixage l et g . Comme il a été mentionné précédemment, le biais des covariances échantillonnées par une matrice diagonale non singulière, aide à stabiliser l'estimateur. Le choix de la forme diagonale comme étant la matrice moyenne des matrices diagonales des valeurs propres des matrices échantillonnées a l'effet de diminuer les valeurs propres importantes et d'augmenter les plus petites. De ce fait, le biais inhérent dans l'estimation des valeurs propres des covariances échantillonnées est contrecarré. Cette forme diagonale est également avantageuse quand les véritables matrices de covariance sont en quelque sorte des multiples de la matrice identité. Cependant, la méthode est coûteuse du point de vue temps de calcul. En outre, l'indice d'optimisation dépend des covariances estimées des autres classes, par conséquent, les mêmes valeurs des paramètres de mixage sont utilisées pour toutes les classes. La même exactitude de classification peut se produire sur une gamme étendue de valeurs de la paire (l, g) [35]. Ceci peut engendrer le problème de sous-estimation de covariance pour les classes ayant un nombre suffisant d'échantillons d'apprentissage.

Une méthode Bayésienne empirique a été suggérée [35] dans laquelle le mélange de covariance a la forme :

$$\hat{S}_i(m) = (1 - \alpha_i(m))S_i + \alpha_i C^*(m) \quad 0 \leq \alpha_i \leq 1 \quad (19)$$

où C^* est une matrice de covariance commune appropriée semblable à C . Ici les matrices de covariance échantillonnées sont modélisées comme étant des conséquences de la distribution *a priori* de Wishart inversée. Le paramètre m est déterminé en maximisant la somme des moyennes des probabilités de vraisemblance des échantillons non utilisés dans le processus de régularisation. Cependant, puisque la matrice de covariance commune C^* et le paramètre de mélange α_i sont choisis dans un contexte Bayésien, cette approche exige que le nombre d'échantillons d'apprentissage soit plus grand que la dimensionalité, ce qui n'est pas toujours le cas en pratique.

Deux autres estimateurs de covariance ont été développés [36]. Ces deux estimateurs consistent à estimer les covariances des classes via des mélanges de la covariance échantillonnée et de la covariance commune. Les fonctions de densité de probabilité *a priori* des matrices de covariance des classes sont estimées à partir des échantillons d'apprentissage et sont utilisées pour estimer la valeur appropriée du paramètre de mixage. Les résultats d'une étude type Monté Carlo ont montré que les deux estimateurs proposés aboutissent dans la plupart des cas à des classifications précises. Un estimateur de covariance qui utilise les estimateurs améliorés des valeurs propres a été développé [37]. Cet estimateur optimise l'estimateur de la covariance échantillonnée en terme d'une minimisation d'une fonction de perte. Deux estimateurs de la matrice de covariance ayant la forme de la covariance échantillonnée plus une matrice définie positive arbitraire ont été proposés [38]. Les deux estimateurs ont été comparés aux estimateurs de forme « plus scalaire » et ont montré leur efficacité en terme d'une minimisation d'une fonction de perte. Un estimateur de l'inverse de la matrice de covariance développé dans [39] rétrécit les valeurs propres vers une valeur commune. Il a été montré que cet estimateur n'est relativement meilleur en terme de minimisation d'une certaine fonction de perte que si l'inverse pondéré de la matrice de covariance échantillonnée est utilisé. Une étude de simulation intéressante de type Monté Carlo comparant quatre méthodes d'estimation de covariance a été effectuée [40]. Les résultats obtenus montrent que lorsque le nombre d'échantillons d'apprentissage est limité, un modèle plus simple pour la matrice de covariance peut mener à une efficacité plus élevée de classification même lorsque les prétentions dans ce modèle étaient incorrectes. Une autre étude de simulation de type Monté Carlo comparant quatre estimateurs de covariance à un estimateur de covariance échantillonnée en minimisant deux fonctions de perte a été effectuée [33]. L'étude a montré que trois méthodes parmi les quatre présentent une nette amélioration par rapport à la covariance échantillonnée.

2. 2. 2. b- Nouveaux estimateurs de covariance :

De nouveaux estimateurs de la matrice de covariance, dans les conditions d'un nombre restreint d'échantillons d'apprentissage, ont été proposés récemment [2], [26], [30], [31], [73], [74]. En général, l'idée de base est de stabiliser la variance de l'estimateur en réduisant l'influence de la présence du terme biaisant via un processus de régularisation. Le principe est de mettre en commun les différents types de covariances des classes en utilisant un modèle de mixage approprié. Selon les critères du choix optimal du paramètre de mixage et du modèle de mixage le plus approprié, l'aspect régularisation diffère d'un estimateur à l'autre.

b-1. Estimateur au sens de maximum de la moyenne des logarithmes LOOL :

L'estimateur proposé ici permet de déterminer la matrice de covariance de chaque classe à partir de trois types différents de mises en commun des covariances échantillonnée et commune de la classe [2], [26]. En général, l'estimateur est non singulier et peut être utilisé même lorsque la covariance échantillonnée ou la covariance commune est singulière. Le choix du mélange des covariances le plus approprié pour la classe est effectué selon la valeur optimale du paramètre de mixage. Un critère probabiliste pour le choix optimal de la valeur du paramètre de mixage a été introduit. Ce critère consiste à maximiser la moyenne des logarithmes des probabilités de vraisemblance des échantillons éliminés pendant le processus de régularisation : “*average log Leave-One-Out Likelihood (LOOL)*”. Afin d'améliorer le temps d'exécution, une implémentation efficace de l'estimateur a été proposée [2], en incorporant des approximations dans les calculs.

L'estimateur en question a la forme suivante :

$$\hat{S}_i(\alpha_i) = \begin{cases} (1 - \alpha_i) \text{diag}(S_i) + \alpha_i S_i & 0 \leq \alpha_i \leq 1 \\ (2 - \alpha_i) S_i + (\alpha_i - 1) C & 1 < \alpha_i \leq 2 \\ (3 - \alpha_i) C + (\alpha_i - 2) \text{diag}(C) & 2 < \alpha_i \leq 3 \end{cases} \quad (20)$$

où S_i est la matrice de covariance échantillonnée et C la matrice de covariance commune définie comme étant la moyenne des matrices de covariance échantillonnées :

$$C = \frac{1}{M} \sum_{i=1}^M S_i \quad (21)$$

La variable α_i est le paramètre de mixage qui détermine l'estimateur sélectionné. Si α_i est nulle, la covariance échantillonnée diagonale est utilisée. Si $\alpha_i = 1$, l'estimateur renvoie la covariance échantillonnée estimée. Si $\alpha_i = 2$, la covariance commune est sélectionnée, et si $\alpha_i = 3$, la covariance commune diagonale est retenue. Pour d'autres valeurs de α_i , l'estimateur retenu est un mélange des covariances. Dans le sous-paragraphe suivant, nous

montrons comment la valeur optimale du paramètre α_i peut être estimée à partir des échantillons d'apprentissage.

- Sélection d'un mélange approprié :

La valeur optimale du paramètre de mixage α_i est choisie de sorte que la moyenne des logarithmes des LOOL des échantillons éliminés pendant le processus de régularisation est maximisée. La technique consiste à enlever un échantillon, estimer la moyenne et la covariance à partir des échantillons restants, puis calculer le logarithme de la probabilité de vraisemblance de l'échantillon enlevé sachant la moyenne et la covariance estimées (LOOL). Le processus est ensuite répété pour chaque échantillon enlevé, et la moyenne des logarithmes des LOOL est calculée. La valeur de α_i qui maximise la moyenne des logarithmes des LOOL est alors choisie et le mélange correspondant est retenu pour l'estimation de la matrice de covariance de la classe considérée.

Soient $m_{i/k}$ la moyenne échantillonnée et $S_{i/k}$ la covariance échantillonnée de la classe i sans l'échantillon k telle que :

$$m_{i/k} = \frac{1}{N_i - 1} \sum_{j=1, j \neq k}^{N_i} x_{i,j} \quad (22)$$

$$S_{i/k} = \frac{1}{N_i - 2} \sum_{j=1, j \neq k}^{N_i} (x_{i,j} - m_{i/k})(x_{i,j} - m_{i/k})^T \quad (23)$$

où la notation $/k$ indique que la quantité est calculée sans l'échantillon k . La covariance commune, sans l'échantillon k de la classe i , est donnée par :

$$C_{i/k} = \left(\frac{1}{M} \sum_{j=1, j \neq i}^M S_j \right) + \frac{1}{M} S_{i/k} \quad (24)$$

L'estimateur pour la classe i , sans l'échantillon k , prend alors la forme :

$$\hat{S}_{i/k}(\alpha_i) = \begin{cases} (1 - \alpha_i) \text{diag}(S_{i/k}) + \alpha_i S_{i/k} & 0 \leq \alpha_i \leq 1 \\ (2 - \alpha_i) S_{i/k} + (\alpha_i - 1) C_{i/k} & 1 < \alpha_i \leq 2 \\ (3 - \alpha_i) C_{i/k} + (\alpha_i - 2) \text{diag}(C_{i/k}) & 2 < \alpha_i \leq 3 \end{cases} \quad (25)$$

L'évaluation de la moyenne des logarithmes des LOOL des échantillons enlevés, x_k , de la classe i est déterminée de la manière suivante :

$$LOOL_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \log(f(x_k | m_{i/k}, \hat{S}_{i/k}(\alpha_i))) \quad (26)$$

Ces calculs sont répétés pour plusieurs valeurs de α_i telles que : $0 \leq \alpha_i \leq 3$, et la valeur de α_i maximisant (26) est alors choisie. Une fois la valeur optimale du paramètre α_i

déterminée, le mélange des covariances approprié dans l'expression (20) est connu et la matrice de covariance de la classe i est estimée en utilisant à la fois tous les échantillons d'apprentissage.

- Implémentation de l'estimateur :

L'implémentation directe de l'estimateur nécessite le calcul de l'inverse et du déterminant de la matrice carrée $C_{i/k}$ à chaque échantillon enlevé. Par conséquent, le temps de calcul devient prohibitif. Cependant, une réduction significative dans le temps de calcul peut être obtenue en introduisant des approximations qui permettent de calculer le déterminant et l'inverse de la matrice $C_{i/k}$ d'une manière plus efficace.

Considérons le cas du mélange de covariance échantillonnée et covariance commune ($1 \leq \alpha \leq 2$). La covariance échantillonnée estimée de la classe i sans l'échantillon k peut être réécrite sous la forme suivante [41] :

$$S_{i/k} = \frac{1}{N_i - 2} \sum_{j=1, j \neq k}^{N_i} (x_{i,j} - m_{i/k})(x_{i,j} - m_{i/k})^T = \frac{N_i - 1}{N_i - 2} \left(S_i - \frac{N_i}{(N_i - 1)^2} v \cdot v^T \right) \quad (27)$$

où $v = x_{i/k} - m_i$. L'estimateur de la covariance commune sans l'échantillon k de la classe i prend alors la forme suivante :

$$\begin{aligned} C_{i/k} &= \frac{1}{L} \sum_{j=1, j \neq i}^L S_j + \frac{1}{L} S_{i/k} = \frac{1}{L} \sum_{j=1}^L S_j - \frac{1}{L} S_i + \frac{1}{L} S_{i/k} \\ &= C - \frac{1}{L} S_i + \frac{N_i - 1}{L(N_i - 2)} \left(S_i - \frac{N_i}{(N_i - 1)^2} v \cdot v^T \right) \\ &= C + \left(\frac{N_i - 1}{L(N_i - 2)} - \frac{1}{L} \right) S_i - \frac{N_i - 1}{L(N_i - 2)} \cdot \frac{N_i}{(N_i - 1)^2} v \cdot v^T \\ &= C + \frac{1}{L(N_i - 2)} S_i - \frac{N_i}{L(N_i - 2)(N_i - 1)} v \cdot v^T \end{aligned} \quad (28)$$

Remplaçant (27) et (28) dans (25), l'estimateur pour $1 < \alpha \leq 2$ devient comme suit :

$$\hat{S}_{i/k}(\alpha) = G - z \cdot z^T \quad (29)$$

où :

$$G = \left(\frac{(2 - \alpha)(N_i - 1)}{(N_i - 2)} + \frac{(\alpha - 1)}{L(N_i - 2)} \right) S_i + (\alpha - 1) \cdot C \quad (30)$$

$$z = \sqrt{K} \cdot v, \quad K = \left(\frac{(2 - \alpha)N_i}{(N_i - 2)(N_i - 1)} + \frac{(\alpha - 1)N_i}{L(N_i - 2)(N_i - 1)} \right) \quad (31)$$

Donc $C_{i/k}(\alpha)^{-1}$ peut être calculé d'une manière efficace en utilisant la formule de Sherman-Morrison-Woodbury [42] :

$$\hat{S}_{i/k}(\mathbf{a}_i)^{-1} = G^{-1} + \frac{G^{-1} \cdot \mathbf{z} \cdot \mathbf{z}^T \cdot G^{-1}}{1 - \mathbf{z}^T \cdot G^{-1} \cdot \mathbf{z}} = G^{-1} + \frac{\mathbf{K}_i \cdot G^{-1} \cdot \mathbf{v} \cdot \mathbf{v}^T \cdot G^{-1}}{1 - \mathbf{K}_i \cdot \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}} \quad (32)$$

Il sera utile de réécrire le terme $x_{i/k} - m_i$ comme suit [41] :

$$x_{i,k} - m_{i/k} = \frac{N_i}{N_i - 1} (x_{i,k} - m_i) = \frac{N_i}{N_i - 1} \mathbf{v} \quad (33)$$

Posons $d = \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}$, le terme quadratique dans la PDF Gaussienne devient alors :

$$\begin{aligned} d_{i/k} &= (\mathbf{x}_{i,k} - \mathbf{m}_{i/k})^T \cdot \hat{S}_{i/k}(\mathbf{a}_i)^{-1} (\mathbf{x}_{i,k} - \mathbf{m}_{i/k}) = \left(\frac{N_i}{N_i - 1} \right)^2 \cdot \mathbf{v}^T \cdot \hat{S}_{i/k}(\mathbf{a}_i)^{-1} \cdot \mathbf{v} \\ &= \left(\frac{N_i}{N_i - 1} \right)^2 \left(\mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v} + \frac{\mathbf{K}_i \cdot \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v} \cdot \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}}{1 - \mathbf{K}_i \cdot \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}} \right) \\ &= \left(\frac{N_i}{N_i - 1} \right)^2 \left(\frac{d}{1 - \mathbf{K}_i \cdot d} \right) \end{aligned} \quad (34)$$

Ainsi, le déterminant peut aussi être calculé d'une manière efficace tel que [41] :

$$|C_{i/k}(\mathbf{a}_i)| = |G - \mathbf{z} \cdot \mathbf{z}^T| = |G| (1 - \mathbf{z}^T \cdot G^{-1} \cdot \mathbf{z}) = |G| (1 - \mathbf{K}_i \cdot \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}) = |G| (1 - \mathbf{K}_i \cdot d) \quad (35)$$

La fonction de logarithme de LOOL pour $1 < \mathbf{a}_i \leq 2$ devient :

$$\begin{aligned} \log(f(x_k | m_{i/k}, C_{i/k}(\mathbf{a}_i))) &= -\frac{p}{2} \log(2 \cdot \mathbf{p}) - \frac{1}{2} \log(|\hat{S}_{i/k}(\mathbf{a}_i)|) - \frac{1}{2} d_{i/k} \\ &= -\frac{p}{2} \log(2 \cdot \mathbf{p}) - \frac{1}{2} \log(|G| (1 - \mathbf{K}_i \cdot d)) - \frac{1}{2} d_{i/k} \\ &= -\frac{p}{2} \log(2 \cdot \mathbf{p}) - \frac{1}{2} \log(|G|) - \frac{1}{2} \log(1 - \mathbf{K}_i \cdot d) \\ &\quad - \frac{1}{2} \left(\frac{N_i}{N_i - 1} \right)^2 \left(\frac{d}{1 - \mathbf{K}_i \cdot d} \right) \end{aligned} \quad (36)$$

Donc, au lieu de calculer l'inverse une matrice carrée ($p \times p$) et de trouver son déterminant à chaque échantillon d'apprentissage enlevé, il est seulement nécessaire de calculer une seule fois l'inverse et le déterminant de la matrice G et d'effectuer un calcul relativement simple de type $d = \mathbf{v}^T \cdot G^{-1} \cdot \mathbf{v}$ à chaque échantillon enlevé.

Considérons maintenant le cas du mélange de covariance échantillonnée diagonale et covariance échantillonné ($0 \leq \mathbf{a}_i \leq 1$). En acceptant l'hypothèse que la covariance échantillonnée diagonale change légèrement quand un échantillon est enlevé :

$$diag(S_i) @ diag(S_{i/k})$$

Une réduction significative dans le calcul peut alors être réalisée. Ainsi, l'estimateur pour $0 \leq \mathbf{a}_i \leq 1$ prend la forme :

$$\begin{aligned}
\hat{S}_{i/k}(\mathbf{a}_i) &= (1 - \mathbf{a}_i) \text{diag}(S_{i/k}) + \mathbf{a}_i S_{i/k} = (1 - \mathbf{a}_i) \text{diag}(S_i) + \mathbf{a}_i S_{i/k} \\
&= (1 - \mathbf{a}_i) \text{diag}(S_i) + \frac{\mathbf{a}_i (N_i - 1)}{(N_i - 2)} \left(S_i - \frac{N_i}{(N_i - 1)^2} \cdot \mathbf{v} \cdot \mathbf{v}^T \right) \\
&= (1 - \mathbf{a}_i) \text{diag}(S_i) + \frac{\mathbf{a}_i (N_i - 1)}{(N_i - 2)} \cdot S_i - \frac{\mathbf{a}_i (N_i - 1) N_i}{(N_i - 2)(N_i - 1)^2} \cdot \mathbf{v} \cdot \mathbf{v}^T \\
&= G_2 - \mathbf{z}_2 \mathbf{z}_2^T
\end{aligned} \tag{37}$$

où :

$$G_2 = (1 - \mathbf{a}_i) \text{diag}(S_i) + \frac{\mathbf{a}_i (N_i - 1)}{(N_i - 2)} \cdot S_i \tag{38}$$

$$\mathbf{K}_2 = \frac{\mathbf{a}_i N_i}{(N_i - 2)(N_i - 1)} \tag{39}$$

$$\mathbf{z}_2 = \sqrt{\mathbf{K}_2} \cdot \mathbf{v} \tag{40}$$

$$d_2 = \mathbf{v}^T \cdot G_2^{-1} \cdot \mathbf{v} \tag{41}$$

En suivant les mêmes étapes de calcul que précédemment, la fonction de logarithme de LOOL prend la forme :

$$\begin{aligned}
\log(f(x_k | m_{i/k}, \hat{S}_{i/k}(\mathbf{a}_i))) &= -\frac{p}{2} \log(2 \cdot p) - \frac{1}{2} \log(|G_2|) - \frac{1}{2} \log(1 - \mathbf{K}_2 d_2) \\
&\quad - \frac{1}{2} \left(\frac{N_i}{N_i - 1} \right)^2 \left(\frac{d_2}{1 - \mathbf{K}_2 d_2} \right)
\end{aligned} \tag{42}$$

Pour le cas du mélange de covariance commune et covariance commune diagonale ($2 < \mathbf{a}_i \leq 3$), les calculs peuvent être simplifiés de la même manière en négligeant les petites variations dans la covariance commune quand un échantillon est enlevé, (c.-à-d., $\text{diag}(C) @ \text{diag}(C_{i/k})$). Ainsi, pour $2 < \mathbf{a}_i \leq 3$, le mélange peut être reformulé tel que :

$$\begin{aligned}
\hat{S}_{i/ki}(\mathbf{a}_i) &= (3 - \mathbf{a}_i) C_{i/k} + (\mathbf{a}_i - 2) \text{diag}(C_{i/k}) \cong (3 - \mathbf{a}_i) C_{i/k} + (\mathbf{a}_i - 2) \text{diag}(C) \\
&= (3 - \mathbf{a}_i) \left(C + \frac{1}{L(N_i - 2)} \cdot S_i - \frac{N_i}{L(N_i - 2)(N_i - 1)} \cdot \mathbf{v} \cdot \mathbf{v}^T \right) + (\mathbf{a}_i - 2) \text{diag}(C) \\
&= G_3 - \mathbf{z}_3 \mathbf{z}_3^T
\end{aligned} \tag{43}$$

où :

$$G_3 = (3 - \mathbf{a}_i) \left(C + \frac{1}{L(N_i - 2)} \cdot S_i \right) + (\mathbf{a}_i - 2) \text{diag}(C) \tag{44}$$

$$\mathbf{K}_3 = \frac{(3 - \mathbf{a}_i) N_i}{L(N_i - 2)(N_i - 1)} \tag{45}$$

$$\mathbf{z}_3 = \sqrt{\mathbf{K}_3} \cdot \mathbf{v} \tag{46}$$

$$d_3 = \mathbf{v}^T \cdot G_3^{-1} \cdot \mathbf{v} \tag{47}$$

En utilisant les mêmes étapes que précédemment, la fonction de logarithme de LOOL est calculée de la manière suivante :

$$\log\left(f\left(x_k \middle| m_{i/k}, \hat{S}_{i/k}(\mathbf{a}_i)\right)\right) = -\frac{p}{2} \log(2.p) - \frac{1}{2} \log(|G_3|) - \frac{1}{2} \log(1 - \mathbf{K}_3 d_3) - \frac{1}{2} \left(\frac{N_i}{N_i - 1}\right)^2 \left(\frac{d_3}{1 - \mathbf{K}_3 d_3}\right) \quad (48)$$

Dans ce nouvel estimateur, l'indice d'optimisation, qui est la moyenne des logarithmes des LOOL, dépend seulement de la covariance estimée de la classe considérée. Par conséquent, chaque classe est traitée indépendamment, ce qui permet d'estimer les paramètres de mixage séparément pour chaque classe. L'avantage est que, lorsque certaines classes ont eu plus d'échantillons d'apprentissage que d'autres, il est plus approprié de leur permettre d'avoir leurs propres paramètres de mixage. Cependant, l'inconvénient est de ne pas être directement lié à l'exactitude de la classification.

Avec les approximations introduites dans les calculs des différents cas de mélange des covariances, le nouvel estimateur est plus rapide que l'estimateur RDA. En fait, RDA nécessite le calcul des valeurs propres et des vecteurs propres d'une matrice carrée ($p \times p$) pour chaque valeur du paramètre I . Cette exigence de calcul n'est pas demandée par le nouvel estimateur. Cependant, sans les approximations, l'estimateur RDA devient plus rapide que le nouvel estimateur.

En utilisant des données multidimensionnelles simulées et une image multispectrale réelle de 191 bandes spectrales, fournies par le système AVIRIS, les résultats publiés dans [2] montrent que lorsque le nouvel estimateur est utilisé, avec ou sans approximations, l'exactitude de la classification est élevée.

Le nouvel estimateur peut être très utile pour un algorithme d'extraction de caractéristiques dont l'exactitude d'estimation de la matrice de covariance est importante. Afin de montrer les avantages du nouvel estimateur pour l'extraction des caractéristiques, des résultats intéressants ont été publiés [2] en utilisant une image multispectrale réelle de 191 bandes spectrales fournie par le système AVIRIS. L'algorithme d'extraction de caractéristiques utilisé est le DBFE "*Decision Boundary-based Feature Extraction*" [29]. L'expérience a montré que l'efficacité de l'algorithme est largement meilleure si les covariances des classes sont estimées en utilisant le nouvel estimateur.

b-2. Estimateur au sens de maximum de la moyenne des logarithmes de BLOOL :

Nous exposons ici un autre estimateur qui utilise la formulation Bayésienne. L'estimateur est une extension des travaux publiés dans [25], [35] et [26]. La première forme de mélanges des covariances est obtenue en supposant que le nombre total des échantillons d'apprentissage est plus grand que la dimensionnalité. Dans ce cas, la matrice de covariance commune est non singulière. L'hypothèse de la distribution Gaussienne des échantillons implique que les matrices de covariance échantillonnées S_i sont mutuellement indépendantes et ayant la distribution de Wishart [32] :

$$S_i \sim W\left(\frac{1}{f_i} \cdot S_i, f_i\right) \quad (49)$$

où $f_i = N_i - 1$, et W représente la distribution de Wishart centrale avec f_i degrés de liberté et S_i matrice paramètre. Supposons que le comportement statistique de chaque S_i a une distribution *a priori* de Wishart inversée de sorte que les matrices S_i soient mutuellement indépendantes avec :

$$S_i \sim W^{-1}((t - p - 1), Y, t) \quad t > p + 1 \quad (50)$$

où W^{-1} est la distribution de Wishart inversée de paramètres Y et t pour p dimension. La moyenne *a priori* Y représente la location centrale de la distribution *a priori* de la matrice S_i , et t contrôle la concentration de la matrice S_i autour de la moyenne Y .

En minimisant l'erreur quadratique, l'estimateur de Bayes de S_i est donné par [35] :

$$\hat{S}_i(Y, t) = \frac{f_i}{f_i + t - p - 1} S_i + \frac{t - p - 1}{f_i + t - p - 1} \cdot Y \quad (51)$$

En posant $\alpha_i = \frac{t - p - 1}{f_i + t - p - 1}$ et Y comme étant la matrice en commune estimée C , S_i peut alors être remplacée par des estimateurs partiellement communs de forme :

$$\hat{S}_i = (1 - \alpha_i) S_i + \alpha_i \cdot C \quad 0 \leq \alpha_i \leq 1 \quad (52)$$

La valeur de t peut aussi être exprimée en termes du paramètre α_i telle que :

$$t = \frac{\alpha_i (f_i - p - 1) + p + 1}{1 - \alpha_i} \quad 0 \leq \alpha_i \leq 1 \quad (53)$$

La covariance commune estimée est alors définie par l'estimateur des moindres carrées généralisé de Y , désigné par $C^*(t)$, pour M classes et pour une valeur de t donnée :

$$C^*(t) = \left(\sum_{i=1}^M \frac{f_i}{f_i + t - p - 1} \right)^{-1} \sum_{i=1}^M \frac{f_i}{f_i + t - p - 1} \cdot S_i \quad (54)$$

Quand le nombre total d'échantillons d'apprentissage est inférieur ou égal à la dimension, la matrice de covariance commune devient instable. Dans ce cas, en biaisant les matrices de covariance échantillonnée et commune estimées par une forme de matrice diagonale, peut éviter le problème de singularité. Le biais de ces deux matrices par leurs propres éléments diagonaux est avantageux lorsque la matrice de covariance de classe est ellipsoïde. L'estimateur de covariance prend alors la forme suivante:

$$\hat{S}_i(\mathbf{a}_i) = \begin{cases} (1 - \mathbf{a}_i)diag(S_i) + \mathbf{a}_i.S_i & 0 \leq \mathbf{a}_i < 1 \\ (2 - \mathbf{a}_i)S_i + (\mathbf{a}_i - 1)C^*(t) & 1 \leq \mathbf{a}_i < 2 \\ (3 - \mathbf{a}_i).C + (\mathbf{a}_i - 2)diag(C) & 2 \leq \mathbf{a}_i \leq 3 \end{cases} \quad (55)$$

La maximisation de la moyenne des logarithmes des probabilités de vraisemblance Bayésiennes des échantillons enlevés pendant le processus de régularisation : “*average log Bayesian Leave-One-Out Likelihood (BLOOL)*” est utilisée comme critère de sélection du modèle de mélange approprié. Pour cela, la valeur du paramètre $\mathbf{a}_i$ est fixée et la moyenne des logarithmes des BLOOL est calculée et comparée pour chaque valeur de $\mathbf{a}_i$. L'implémentation directe de la fonction BLOOL pour chaque classe avec N_i échantillons nécessitera le calcul de N_i inverses de matrice et N_i déterminants à chaque valeur de $\mathbf{a}_i$. Cependant, une implémentation plus efficace a été proposée en utilisant un algorithme d'approximation des calculs [30], qui ne sera pas donné ici.

Puisque ce nouvel estimateur est une extension des estimateurs RDA, LOOL, et l'estimateur Bayésien empirique de covariance, il bénéficie des avantages inhérents dans ces trois estimateurs. Ainsi, tout comme l'estimateur basé sur le critère LOOL, l'estimateur basé sur le critère BLOOL est tout à fait flexible vis-à-vis du nombre d'échantillons d'apprentissage en traitant des ensembles d'échantillons d'apprentissage de tailles différentes, le paramètre de mélange est adapté pour chaque classe. En outre, avec l'approximation sur les matrices diagonales, l'estimateur BLOOL est plus efficace en temps de calcul que l'estimateur RDA et l'estimateur Bayésien empirique de covariance. De plus, tout comme l'estimateur Bayésien empirique de covariance, il peut s'adapter à la variabilité des matrices de covariance échantillonnées.

Cependant, l'estimateur BLOOL souffre aussi des inconvénients inhérents dans RDA, LOOL, et l'estimateur Bayésien empirique de covariance. Tout d'abord, il a l'inconvénient de n'avoir aucune relation directe avec l'exactitude de classification. Quoique l'instabilité des estimateurs de covariance dans le cas d'un petit nombre d'échantillons d'apprentissage puisse être réduite en utilisant l'une parmi les approches de mélange des covariances citées ci-dessus, le problème est que le degré d'amélioration est limité. En fait, les estimateurs des matrices de covariance et les paramètres de mélange utilisés dans les mélanges des covariances sont basés sur des nombres limités d'échantillons d'apprentissage. En particulier, quand les nombres d'échantillons d'apprentissage sont assez petits, les matrices de covariance ainsi estimées sont adaptées aux nombres d'échantillons seulement et ne peuvent pas être représentatives des vraies statistiques de l'ensemble entier des données.

En utilisant une image multispectrale réelle de 200 bandes spectrales fournies par le système AVIRIS, une étude comparative entre ce nouvel estimateur, qui est basé sur le critère BLOOL, et le nouvel estimateur basé sur le critère LOOL donnée précédemment a été effectuée dans [31]. L'étude consiste à classer les 200 bandes spectrales par plusieurs

classifieurs qui utilisent les statistiques d'ordre un et deux des classes, en utilisant les deux nouveaux estimateurs. Les résultats obtenus ont montré que l'estimateur basé sur le critère BLOOL a des performances meilleures quand les échantillons limités reflètent les vraies distributions *a priori* des classes. Cependant, comme le cas de l'estimateur basé sur le critère LOOL, l'estimateur basé sur le critère BLOOL a l'inconvénient de ne pas être directement lié à la séparabilité des classes, et donc à l'exactitude de la classification. Par conséquent, des nouveaux critères Bayésiens dérivés de la séparabilité des classes ont été recommandés pour les futurs recherches [30], [31].

b-3. Estimateurs types LOOL mélangés :

Récemment, Kuo et Landgrebe [73] ont proposé une paire de nouveaux estimateurs régularisés améliorés de covariance appelées LOOL mélangés "*Mixed-LOOLs (MLOOLs)*" afin d'éviter l'erreur excessive de l'estimateur de covariance.

LOOL et BLOOL sont des combinaisons linéaires de deux matrices parmi trois matrices différentes utilisées dans les deux estimateurs (voir expressions (20), (25) et (55)). Il y a des situations où BLOOL est meilleur que LOOL, comme il y a d'autres situations où LOOL est le meilleur [31]. La différence entre LOOL et BLOOL réside dans les types de matrices utilisées pour formuler l'estimateur de covariance régularisé. Donc, l'utilisation seulement de quelques matrices parmi les six matrices possibles n'aboutit pas dans toutes les situations à de bons résultats. L'idée de base des estimateurs MLOOLs est d'utiliser les six matrices à la fois afin d'exploiter les avantages de LOOL et de BLOOL. Ainsi, le premier estimateur MLOOL1 a la forme suivante :

$$\begin{aligned}\hat{S}_i(\mathbf{q}_i) &= \hat{S}_i(a_i, b, c_i, d_i, e, f_i) \\ &= a_i \cdot \frac{\text{tr}(S_i)}{p} \cdot I + b \cdot \text{diag}(S_i) + c_i \cdot S_i + d_i \cdot \frac{\text{tr}(C)}{p} \cdot I + e \cdot \text{diag}(C) + f_i \cdot C\end{aligned}\quad (56)$$

où $a_i + b + c_i + d_i + e + f_i = 1$ et $i = 1, 2, \dots, M$. Les paramètres de mélange sont déterminés en maximisant LOOL de chaque classe :

$$LOOL_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \log \left(f(x_k | m_{i/k}, \hat{S}_{i/k}(\mathbf{q}_i)) \right) \quad (57)$$

Cependant, l'estimateur MLOOL1 est intensif du point de vue calcul [73], un autre estimateur plus simplifié sera plus pratique. Puisque la valeur optimale du paramètre $\mathbf{a}$ n'est pas connue et dépend des valeurs des LOOL calculées pour chaque échantillon enlevé, alors pour une valeur fixe $\mathbf{a}$, faire un choix d'une matrice parmi les six peut alors réduire le temps de calcul. Ainsi, le second estimateur type mélange de LOOL, noté MLOOL2, est proposé sous la forme suivante [73] :

$$\hat{S}_i(\mathbf{a}_i) = \mathbf{a}_i \cdot A + (1 - \mathbf{a}_i)B \quad (58)$$

où A peut prendre l'une des formes suivantes : $\frac{tr(S_i)}{p}.I$, $diag(S_i)$, S_i , $\frac{tr(C)}{p}.I$, $diag(C)$, ou C et B peut prendre l'une des deux formes suivantes : S_i , ou $diag(C)$. La valeur du paramètre α est choisie très proche de 1. Le choix de B comme étant égale à S_i , ou $diag(C)$ est justifié par le fait que si le nombre d'échantillons d'apprentissage est important, S_i sera le meilleur choix. Si le nombre est insuffisant alors la matrice de covariance commune est choisie. Mais cette matrice peut être singulière. Pour réduire l'erreur de l'estimation et éviter la singularité, $diag(S)$ sera le bon choix. Le critère de sélection est la fonction LOOL pour une valeur du paramètre α fixe :

$$LOOL_i = \frac{1}{N_i} \sum_{k=1}^{N_i} \log \left(f(x_k | m_{i/k}, \hat{S}_{i/k}(\alpha)) \right) \quad (59)$$

La mise en œuvre de l'estimateur MLOOL2 pour chaque classe consiste à calculer des LOOL des 12 combinaisons des estimateurs de covariance et choisir LOOL dont la valeur est la plus élevée. Cet estimateur amélioré est inversible pour chaque classe. Il présente des avantages par rapport aux estimateurs LOOL [2], [26] et BLOOL [30] du point de vue robustesse et temps de calcul. Les résultats publiés dans [73], utilisant des données simulées et des images multispectrales réelles de 191 bandes fournies par le système AVIRIS, ont montré en général que l'estimateur MLOOL2 donne de meilleurs résultats par rapport à l'estimateur LOOL. En particulier, lorsque la dimensionnalité de l'image multispectrale augmente ou lorsque le nombre d'échantillons d'apprentissage pour chaque classe est différent. Dans les situations où l'estimateur BLOOL donne de bons résultats par rapport à l'estimateur LOOL, les estimateurs type MLOOL donnent des résultats comparables ou meilleurs. Ceci est justifié par le fait que les concepts des estimateurs MLOOL ont été faits pour retenir les avantages des estimateurs LOOL et BLOOL et éviter leurs inconvénients. Ainsi les estimateurs MLOOL sont plus stables par rapport aux estimateurs LOOL et BLOOL. Cependant, le temps de calcul de l'estimateur MLOOL1 est supérieur à ceux de LOOL et BLOOL.

b-4. Estimateurs adaptatifs des covariances combinées :

Les méthodes de régularisation essayent de réduire la variance de l'estimateur de covariance en le biaisant vers des valeurs qui sont considérées physiquement plausibles [25]. Par conséquent, la variance est réduite en limitant le nombre de paramètres requis pour être estimé via la valeur de biais rajouté. L'ampleur de cette compensation de biais-variance est contrôlée par un ou plusieurs paramètres de régularisation. Dans cette optique, une nouvelle méthode qui ajuste les estimateurs régularisés de covariance via la précision d'une classification adaptative est proposée dans [74]. Selon la façon de choisir les matrices de covariance du mélange et les paramètres de régularisation, de nouveaux estimateurs adaptatifs de covariance sont présentés. Les paramètres de régularisation et les

matrices de covariance utilisées dans le mélange sont déterminés en se basant sur les échantillons d'apprentissage et les échantillons semi-classifiés. Les échantillons semi-classifiés sont les sorties du classifieur pour un type de paramètres de régularisation et matrices de mélange. Ces paramètres et ces matrices sont mises à jour plusieurs fois jusqu'à ce que l'exactitude de classification la plus élevée soit atteinte. Les résultats obtenus, en utilisant des données simulées, montrent que pour un nombre limité d'échantillons, les estimateurs proposés réalisent les mêmes performances de classification que celle obtenues si un nombre plus important d'échantillons d'apprentissage est utilisé. Dans le cas d'images hyper-spectrales réelles, les résultats obtenus indiquent que les estimateurs proposés peuvent améliorer la précision de classification de manière significative en utilisant un nombre restreint d'échantillons d'apprentissage. Nous détaillons dans ce qui suit cette classe des nouveaux estimateurs de covariance.

Initialement, les matrices de covariance régularisées sont déterminées en utilisant les échantillons d'apprentissage. Elles sont consécutivement mises à jour en utilisant les échantillons d'apprentissage et les échantillons semi-classifiés jusqu'à ce que la convergence soit atteinte. Selon le type d'estimateur de covariance utilisé, les estimateurs adaptatifs ont plusieurs formes. Les étapes principales de l'estimation sont les suivantes :

(1) Calcul des facteurs de pondération tels que :

$$w_{ij}^c = \frac{1}{1 + \sum_{k=1, k \neq j}^M \exp\left(-\frac{1}{2} \cdot (d_k(x_{ij}) - d_l(x_{ij}))\right)} \quad (60)$$

où $d_i(x)$ est la distance qui sépare le pixel x par rapport à la moyenne de la classe i . Elle dépend du type de classifieur utilisé.

(2) Estimation des vecteurs moyenne (m_i^+) et les matrices de covariance échantillonnées (S_i^+) et communes (C^+) :

$$m_i^+ = \frac{\sum_{j=1}^{m_i} y_{ij} + \sum_{j=1}^{n_i} w_{ij}^c \cdot x_{ij}^c}{m_i + \sum_{j=1}^{n_i} w_{ij}^c} \quad (61)$$

$$S_i^+ = \frac{\sum_{j=1}^{m_i} (y_{ij} - m_i^+)(y_{ij} - m_i^+)^T + \sum_{j=1}^{n_i} w_{ij}^c (x_{ij}^c - m_i^+)(x_{ij}^c - m_i^+)^T}{m_i + \sum_{j=1}^{n_i} w_{ij}^c} \quad (62)$$

$$C^+ = \frac{1}{M} \cdot \sum_{i=1}^M S_i^+ \quad (63)$$

où $y = (y_{i1}, y_{i2}, \dots, y_{im})$ dénote les échantillons d'apprentissage de la classe i et $x = (x_{i1}^c, x_{i2}^c, \dots, x_{im}^c)$ les échantillons semi-classifiés qui ont été classifiés appartenant à la classe i .

- (3) Estimation du mélange de covariances régularisé. Cette étape dépend du type d'estimateur de mélange de covariance utilisé (LOOL ou BLOOL).
- (4) Effectuer une classification basée sur la règle de maximum de vraisemblances.
- (5) Répéter les étapes (1), (2), (3), et (4) jusqu'à la convergence soit atteinte.

Puisque les échantillons semi-classifiés sont utilisés avec les échantillons d'apprentissage, les critères LOOL et BLOOL pour la sélection des paramètres de mélange sont modifiés dans l'étape (3). En d'autres termes, les paramètres de mélange sont choisis de sorte que le meilleur ajustement par rapport aux échantillons d'apprentissage et semi-classifiés soit réalisé, ce qui implique également qu'une meilleure exactitude de classification est réalisée. Cependant, pour chaque classe i avec m_i échantillons d'apprentissage et n_i échantillons semi-classifiés, l'implémentation directe des estimateurs adaptatifs en utilisant les critères LOOL et BLOOL nécessite le calcul de $(m_i + n_i)$ matrices inverses et déterminants pour chaque valeur de paramètre de mélange. Une implémentation plus efficace peut être obtenue [74] en utilisant les approximations dans le calcul des matrices inverses et de ses déterminants [26].

3. Caractérisation Statistique et Modélisation d'Images Radars

Le processus de formation de l'image radar est très compliqué par rapport à celui de l'image optique. L'image radar SAR a des propriétés particulières qui nécessitent de prendre en considération d'autres aspects pour décrire les comportements des signaux radars réfléchis. En fait, dans la gamme des micro-ondes, la rugosité, les propriétés électriques, et la forme géométrique de l'objet terrestre influent sur les rayonnements réfléchis [43]. L'image SAR est sévèrement dégradée par le speckle [44]-[46]. Le speckle a les caractéristiques d'un bruit multiplicatif et présente des défis à la plupart des modèles proposés pour l'analyse et classification des images SAR.

Dans la littérature, on trouve deux approches différentes pour la description des comportements des signaux radars réfléchis par les objets terrestres. Les approches déterministes basées sur le principe du problème inverse, en déduisant les relations analytiques des coefficients de dispersion à partir des connaissances *a priori* des propriétés électromagnétiques de l'objet terrestre. Les approches statistiques qui consistent à

caractériser et modéliser les comportements des signaux réfléchis directement à partir des images SAR fournies par les systèmes d'acquisition.

Dans les approches déterministes, plusieurs modèles sophistiqués ont été développés [43], [46]-[51]. Certains d'entre eux nécessitent la définition de toutes les fonctions requises pour reconstruire la scène imagée à partir des signaux reçus. De tels modèles, appelés déterministes, sont basés sur les équations de propagation des ondes émises et réfléchies comme point de départ pour déterminer les relations analytiques des coefficients de dispersion, et ceci à partir des connaissances *a priori* des propriétés électromagnétiques de l'objet terrestre [43], [46]. Cependant, l'élaboration de ces modèles demande des simplifications diverses afin de réduire le nombre important de paramètres d'entrée, ce qui limite leur efficacité dans l'analyse et classification d'images SAR. En d'autres termes, ces modèles ont été proposés pour décrire le comportement des signaux réponses à partir des effets géométriques et physiques, et non pas pour analyser et traiter l'image SAR. Une autre approche déterministe développée consiste à dériver les propriétés de la scène observée sachant les images SAR, les données physiques, et autres. C'est un problème inverse classique mal posé dans le sens où il approche la situation en intégrant dans un seul modèle l'information portée par l'image SAR avec les connaissances *a priori* des propriétés physiques des objets terrestres et les conditions atmosphériques et autres à l'instant de la prise de vue. Le modèle ainsi dérivé peut être de type phénoménologique basé sur la théorie de dispersion et les effets physiques et atmosphériques. Cependant, l'approche de la théorie de dispersion nécessite l'utilisation des techniques de simulation pour décrire le problème en termes de formes géométriques et tailles des objets terrestres, ainsi que la connaissance détaillée des paramètres du modèle [47]. Dans la pratique, cette approche est très fastidieuse, coûteuse, et souvent impossible.

La caractérisation statistique et la modélisation des images SAR permettent de développer des algorithmes d'analyse et de classification ainsi que d'élaborer des filtres spécialisés pour la réduction du speckle. Durant cette dernière décennie, plusieurs études ont été effectuées afin de relier les propriétés physiques et le comportement statistique des images SAR. Et afin d'établir cette relation, des hypothèses et des distributions statistiques ont été supposées. Les différents modèles probabilistes proposés sont basés sur des considérations suivant que l'image SAR soit une image amplitude ou image intensité, si elle est multivues ou non, et si la distribution est Gaussienne, Gaussienne circulaire, ou Gaussienne mixte, etc. Pour les différents cas de l'image SAR, il a été montré que les distributions Gaussiennes ne sont pas appropriées pour décrire les réponses radars réfléchies et ne reflètent pas la réalité physique du terrain. Par conséquent, les chercheurs se dirigent vers la considération d'autres modèles probabilistes pour la caractérisation statistiques des images SAR [8], [9], [52], [20], [19]. Plusieurs modèles non-Gaussiens [53] ont été utilisés et testés. En premier lieu, la distribution de Rayleigh a été utilisée pour

modéliser l'image SAR d'amplitude. La distribution est conforme pour les régions homogènes et pour les images de grande résolution spatiale. Cependant, elle n'est pas souvent valable pour les régions hétérogènes et pour les images de résolution spatiale plus ou moins fine [9]. D'autres distributions ont alors été testées telles que les distributions b [7], [54], Gamma [8], Lognormal [6], Weibull [5], la distribution K [10]-[11], [17]-[22], [55]-[58], et la distribution de Wishart [12]-[16]. Parmi toutes ces distributions, seule la distribution K a suscité beaucoup plus d'attention dans la communauté. La distribution de Wishart a été récemment proposée mais elle n'est pas très utilisée. Pour cette raison, nous allons consacrer les deux sous paragraphes suivants à ces deux distributions. Dans le premier sous paragraphe, la distribution de Wishart est considérée. Nous essayons de justifier pourquoi son utilisation comme modèle pour les images SAR est restrictive. Dans le sous paragraphe suivant, la distribution K est considérée [11]. Les autres distributions ne seront pas abordées, mais elles seront comparées à la distribution K [11].

3. 1. Modélisation par la Distribution Complexe de Wishart

Le principe de fonctionnement des radars SAR, traditionnels et polarimétriques, ainsi que la nature et l'origine du speckle et l'introduction des paramètres géométriques propres à la prise de vue de l'image SAR sont donnés dans notre publication [11] (voir la Partie C pour la version intégrale de la publication). Afin de rendre plus facile la compréhension de ce paragraphe, nous rappelons dans les deux sous paragraphes suivants la formulation mathématique des images POLSAR en terme de représentation des cibles distribuées (ou étendues). La modélisation des images POLSAR par la distribution complexe de Wishart est donnée par la suite.

3. 1. 1. Formulation des images POLSAR

Le radar conventionnel utilise la même configuration de polarisation à la fois pour l'émission et la réception du signal électromagnétique. Ce système ne permet donc pas une caractérisation complète des mécanismes d'interaction de l'onde électromagnétique émise avec la surface. Les propriétés de dépolarisation des cibles naturelles ne peuvent être étudiées. En revanche, le radar polarimétrique acquiert toute l'information contenue dans le phénomène de rétro-diffusion, en mesurant, pour chaque élément de la scène, l'ensemble des quatre polarisations de la réponse, en amplitude et en phase relative. Ainsi, le radar AirSAR du JPL-NASA, par exemple, possède deux antennes émettrices pour chacune des polarisations horizontale (H) et verticale (V), ainsi que deux récepteurs H et V . Le radar émet alternativement des impulsions en polarisations H et V , tandis que l'onde rétro-diffusée est reçue simultanément en polarisations H et V . L'ouverture synthétique est

réalisée par intégration cohérente, pour chacune des polarisations HH , HV , VH , et VV , sur plusieurs impulsions. Toute polarisation s'écrivant comme une combinaison linéaire des deux précédentes (H et V). Ainsi, pour un angle d'incidence donné, le radar polarimétrique fournit les images POLSAR sous forme d'une matrice carrée (2×2) complexe (amplitude et phase) appelée matrice de rétro-diffusion S telle que [49], [66], [43] :

$$S = \begin{pmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{pmatrix} \quad (64)$$

où S_{ij} est l'élément de rétro-diffusion de polarisation de type transmission i et réception j , sachant que i et j peuvent être H ou V . Théoriquement, les composantes S_{HH} et S_{VV} sont corrélées. Tandis que pour les composantes co-polaires et contra-polaires, c-a-d, S_{HH} (ou S_{VV}) et l'élément S_{ij} (ou S_{ji}), avec $i = H$ et $j = V$, la corrélation est nulle. Cependant, les images POLSAR fournies indiquent que cette corrélation n'est pas tout à fait nulle, mais elle est égale à une certaine valeur faible. Ceci est probablement dû au bruit thermique négligé dans le concept théorique de la matrice de rétro-diffusion [69].

Il existe d'autres représentations matricielles des caractéristiques des mécanismes de rétro-diffusion. La représentation par la matrice de Mueller est particulièrement importante car elle permet de traiter de façon simple aussi bien les cibles ponctuelles que les cibles distribuées. Cette matrice est carrée, réelle, de dimension 4×4 et peut être reliée à la matrice de rétro-diffusion [43]. La représentation par la matrice de covariance polarimétrique, $C = S.S^{*T}$, présente les mêmes avantages que celle de la matrice de Mueller. Un autre type de représentation est la représentation par la matrice de cohérence, obtenue à partir de la matrice de covariance par une simple transformation linéaire. C'est à partir de ces deux dernières représentations que la modélisation des images POLSAR par la distribution complexe de Wishart a été déduite. Nous allons décrire dans ce qui suit la caractérisation des mécanismes de rétro-diffusion des cibles distribuées par la matrice de cohérence. Ensuite, la déduction de la distribution complexe de Wishart comme modèle pour les images POLSAR est donnée.

3. 1. 2. Caractérisation des cibles distribuées par la matrice de cohérence

Dans la rétro-diffusion d'une onde électromagnétique par une cible, les champs électriques des ondes transmises et reçues sont exprimées dans les systèmes locaux centrés sur les antennes de transmission et de réception [67], tandis que le système global est centré sur la cible visée. Dans le cas du radar mono-statique, c-a-d, des antennes situées en un même lieu, la convention BSA "*Back Scattering Alignment*" fait coïncider les deux repères. Par conséquent, la vérification du principe de réciprocité implique que la rétro-diffusion est

réciroque, et on a alors $S_{HV} = S_{VH}$ [67], [43]. Dans ce cas, l'information de la rétro-diffusion polarimétrique peut être représentée par un vecteur polarimétrique complexe h :

$$h = \begin{pmatrix} S_{HH} & \sqrt{2}.S_{HV} & S_{VV} \end{pmatrix}^T \quad (65)$$

Le terme $\sqrt{2}$ dans la deuxième composante du vecteur sert à assurer la cohérence avec le calcul de la puissance totale (*span* en anglais) [70]. Dans le cas d'une rétro-diffusion non-réciroque, la dimension du vecteur h est égale à 4. Quand le radar éclaire une zone d'une surface contenant plusieurs cibles élémentaires (supposées distribuées aléatoirement), l'onde rétro-diffusée est la somme cohérente des ondes rétro-diffusées par les cibles élémentaires [11]. Cette onde est une variable aléatoire au cours du temps. La polarisation de l'onde a changé et les ondes transmises et réfléchies sont alors reliées par la matrice de rétro-diffusion complexe S . Chaque cible est donc caractérisée par sa propre matrice de rétro-diffusion. Le vecteur h peut être modélisé par une distribution Gaussienne multivariée complexe [53]:

$$P(h) = \frac{1}{\pi^3 |C|} \cdot \exp(-h^{*T} \cdot C^{-1} \cdot h) \quad (66)$$

où $*$ signifie le complexe conjugué, C est la matrice de covariance complexe polarimétrique, et $|C|$ son déterminant. La matrice C est une matrice Hermitienne, c-à-d $C = C^{*T}$, obtenue à partir du vecteur h telle que:

$$C = E(h \cdot h^{*T}) = \begin{pmatrix} |S_{HH}|^2 & \sqrt{2}.S_{HH}.S_{HV}^* & S_{HH}.S_{VV}^* \\ \sqrt{2}.S_{HV}.S_{HH}^* & 2|S_{HV}|^2 & \sqrt{2}.S_{HV}.S_{VV}^* \\ S_{VV}.S_{HH}^* & \sqrt{2}.S_{VV}.S_{HV}^* & |S_{VV}|^2 \end{pmatrix} \quad (67)$$

A partir de (67), la puissance totale (*span*) est exprimée comme étant :

$$span = h^{*T} \cdot h = |S_{HH}|^2 + 2|S_{HV}|^2 + |S_{VV}|^2 \quad (68)$$

D'après (65) et (68), nous remarquons que le module du vecteur h est égal à la puissance totale (*span*). Les parties réelles et imaginaires des éléments complexes du vecteur h sont supposées avoir une distribution Gaussienne circulaire [72].

Les images POLSAR monovues sont souvent transformées en images multivues afin de réduire le speckle et compresser le volume de données. L'information polarimétrique est alors contenue dans la matrice de covariance $\langle C_i \rangle$ de chaque pixel i de l'image multivues dont l'expression est donnée par:

$$\langle C_i \rangle = \frac{1}{L} \sum_{j=1}^L C_j = \begin{pmatrix} \langle |S_{i,HH}|^2 \rangle & \langle \sqrt{2}.S_{i,HH}.S_{i,HV}^* \rangle & \langle S_{i,HH}.S_{i,VV}^* \rangle \\ \langle \sqrt{2}.S_{i,HV}.S_{i,HH}^* \rangle & \langle 2|S_{i,HV}|^2 \rangle & \langle \sqrt{2}.S_{i,HV}.S_{i,VV}^* \rangle \\ \langle S_{i,VV}.S_{i,HH}^* \rangle & \langle \sqrt{2}.S_{i,VV}.S_{i,HV}^* \rangle & \langle |S_{i,VV}|^2 \rangle \end{pmatrix} \quad (69)$$

où L est le nombre de vues (*look* dans la terminologie anglaise) et C_j est la matrice de covariance complexe polarimétrique du pixel j dans l'image monovue. La matrice de cohérence du pixel i dans l'image multivues, notée $\langle T_i \rangle$, est obtenue à partir de la matrice de covariance $\langle C_i \rangle$ par la transformation linéaire suivante [14]:

$$\langle T_i \rangle = E \langle C_i \rangle E^T \quad \text{où} \quad E = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & -1 \\ 0 & \sqrt{2} & 0 \end{pmatrix} \quad (70)$$

La représentation de la matrice de cohérence a l'avantage sur la matrice de covariance d'être relié aux mécanismes de la rétro-diffusion physiques de la cible. Notons qu'il y a une autre approche pour obtenir la matrice de cohérence directement à partir de la matrice de rétro-diffusion sans passer par la matrice de covariance [15]. Cependant, cette approche demande trop de calculs algébriques.

3. 1. 3. Distribution de Wishart comme modèle pour les images POLSAR

Pour commencer, nous donnons la définition mathématique de la distribution de Wishart ainsi que sa signification du point de vue statistique [32].

Si un processus aléatoire $\mathbf{X}$ constitué des variables aléatoires X_i , avec $i = 1, 2, \dots, n$, a une distribution Gaussienne multidimensionnelle, avec un vecteur moyenne nul $m = 0$ et une matrice de covariance $\mathbf{S}$, et $\mathbf{Y}$ dénote la matrice ($n \times p$) composée de p vecteurs lignes X_i , alors la matrice carrée $\mathbf{Y} = \mathbf{Y}^T \cdot \mathbf{Y}$ a la distribution de Wishart de matrice paramètre $\mathbf{S}$ et degrés de liberté n . Cette distribution est notée par :

$$\mathbf{Y} \sim W_p(\mathbf{S}, n) \quad (71)$$

La distribution de Wishart est utilisée pour décrire le comportement statistique de la matrice de covariance échantillonnée d'un processus aléatoire Gaussien. En fait, si la distribution de Chi-deux est utilisée pour faire l'inférence dans le cas d'une variable aléatoire Gaussienne mono-dimensionnelle, la distribution de Wishart peut être utilisée pour le même but dans le cas d'un processus aléatoire Gaussien multidimensionnel. Elle possède deux propriétés importantes qui sont la pondération et l'aditivité. Si la matrice

aléatoire Y est pondérée par une matrice carrée de dimension p , soit F , on a alors : $F^T.Y.F \sim W_p(F^T.S.F, n)$. Si on a deux matrices aléatoires Gaussiennes, Y_1 et Y_2 , telles que : $Y_1 \sim W_p(S, n_1)$ et $Y_2 \sim W_p(S, n_2)$, avec $Cov(Y_1, Y_2) = 0$, alors $Y_1 + Y_2 \sim W_p(S, n_1 + n_2)$. Ces deux propriétés seront utilisées dans l'application de cette distribution en modélisation d'images POLSAR.

Si on considère la variable aléatoire X_i du processus aléatoire $\mathbf{X}$ comme étant le vecteur polarimétrique complexe h_i au niveau du pixel i d'une image monovues, et si on considère l'ensemble de toutes les variables aléatoires du processus $\mathbf{X}$ comme étant l'ensemble de tous les vecteurs polarimétriques complexes des pixels utilisés pour obtenir l'image multivues, par identification on a alors $n = L$ et la matrice Y est formée par les vecteurs complexes h_i . Cette matrice a la dimension $(L \times q)$, où q est la dimension du vecteur h , et caractérise un pixel i dans l'image multivues, qu'on note alors Y_i . Par identification, la matrice carrée $\mathbf{Y}_i = Y_i^T.Y_i$ est donc la matrice de covariance $\langle C_i \rangle$ d'un pixel i de l'image multivues. Puisque les vecteurs h_i sont supposés complexes Gaussiens [72], alors, d'après la définition de la distribution de Wishart donnée ci-dessus, les statistiques de la matrice de covariance complexe polarimétrique $\langle C_i \rangle$ (ou $\mathbf{Y}_i$) possèdent la distribution complexe de Wishart [53], [69]. Puisque la matrice de covariance $\langle C_i \rangle$ peut être convertie en une matrice de cohérence $\langle T_i \rangle$ par une simple transformation linéaire (voir expression (70)) et donc, d'après la propriété de pondération de la distribution de Wishart, les statistiques de la matrice de cohérence ont aussi la distribution complexe de Wishart.

Dans la littérature, la matrice de cohérence a été utilisée au lieu de la matrice de covariance dans la déduction du modèle de Wishart pour les images POLSAR. Ceci est dû au fait que la représentation par la matrice de cohérence a l'avantage sur la matrice de covariance d'être relié aux mécanismes de la rétro-diffusion physiques de la cible. Ainsi, la distribution complexe de Wishart de la matrice de cohérence $\langle T_i \rangle$ d'un pixel i d'une image multivues (L vues) sachant la classe w_m est donnée par [53] :

$$P_{T_i}^{(L)}(\langle T_i \rangle | w_m) = \frac{L^q |\langle T_i \rangle|^{L-q} \cdot \exp(-L \cdot \text{Tr}(V_m^{-1} \langle T_i \rangle))}{K(L, q) |V_m|^L} \quad (72)$$

où q est la dimension du vecteur du cible h , avec $q = 3$ dans le cas de la réciprocité (c.-à-d., $S_{HV} = S_{VH}$) et $q = 4$ dans le cas de la non-réciprocité. Tr signifie la trace de la matrice. V_m est la matrice moyenne des matrices de cohérence des pixels appartenant à la classe w_m :

$$V_m = E(\langle T_i \rangle | \langle T_i \rangle \in w_m) = \frac{1}{N_m} \sum_{i=1}^{N_m} \langle T_i \rangle \quad (73)$$

où N_m est le nombre de pixels de la classe w_m . $K(L, q)$ est le facteur de normalisation donné par [53], [69] :

$$K(L, q) = p^{(1/2)q(q-1)} \cdot G(L) \cdot G(L - q + 1) \quad (74)$$

où $G(\cdot)$ est la fonction de Gamma. Le modèle (72) a été utilisé dans l'algorithme de classification Bayésienne au sens de maximum de vraisemblances proposé dans [13]. La règle de décision du classifieur n'est rien que la forme logarithmique de la règle Bayésienne de maximum de vraisemblances [4] avec comme PDF la distribution complexe de Wishart donnée dans l'expression (72). Une mesure probabiliste de la distance entre la matrice de cohérence d'un pixel quelconque, $\langle T_i \rangle$, et la matrice moyenne des matrices de cohérence de la classe candidate w_m , V_m , est alors obtenue:

$$d(\langle T_i \rangle, V_m) = L(\log(|V_m|) + \text{Tr}(V_m^{-1} \langle T_i \rangle)) - \log(P(w_m)) \quad (75)$$

où $P(w_m)$ est la probabilité *a priori* de la classe w_m . En général, les classes sont supposées équiprobables et l'expression (75) devient :

$$d(\langle T_i \rangle, V_m) = \log(|V_m|) + \text{Tr}(V_m^{-1} \langle T_i \rangle) \quad (76)$$

Le pixel en considération est affecté à la classe w_m si et seulement si :

$$d(\langle T_i \rangle, V_m) < d(\langle T_i \rangle, V_j) \quad \text{pour tout } j \neq m \quad (77)$$

3. 1. 4. Utilisation restrictive de la distribution de Wishart

Dans la communauté de télédétection radar, on ne trouve pas beaucoup de chercheurs qui utilisent la distribution de Wishart comme modèle pour les images SAR. Jusqu'à présent, la plupart des travaux publiés dans ce contexte restent ceux de Cloude, Pottier, Grunes et Lee [12]-[16], [71] qui, en exploitant les résultats théoriques du Goodman [53], sont les premiers qui ont proposé ce modèle. Nous essayons dans ce sous-paragraphe de justifier en plus cette utilisation restrictive de la distribution de Wishart.

Tout d'abord, on note que la distribution de Wishart décrit le comportement statistique de la matrice de covariance polarimétrique (ou matrice de cohérence) d'un pixel d'une image POLSAR multivues appartenant à une classe et non pas le comportement statistique de la classe elle-même. Ceci a engendré une certaine ambiguïté dans l'appellation modèle statistique pour la classe qui d'habitude caractérise le comportement statistique de toute la classe. La déduction de la distribution de Wishart comme modèle pour les cibles distribuées « étendues » des images POLSAR multivues a été obtenue localement, c-a-d,

au niveau de chaque pixel, l'inférence statistique de la matrice de cohérence du pixel est faite. La question qui se pose est la suivante : Pour une classe donnée, les matrices de cohérence des pixels de la classe, ont-elles le même comportement statistique, c-a-d, suivent-elles la distribution de Wishart ? Certes la réponse est non puisque le phénomène speckle n'a pas été pris en considération dans la dérivation du modèle. Le speckle est un phénomène qui se produit chaque fois qu'il y a des interférences entre les réponses réfléchies [11]. La rugosité de surface est suffisante pour engendrer le phénomène speckle. En outre, on note que pour des cibles ponctuelles ainsi que pour des images monovues, la distribution complexe de Wishart n'est plus valable [68].

L'absence d'un fondement théorique convaincant et purement statistique sur la convenance de la distribution de Wishart comme modèle pour les images SAR, est une autre raison justifiant l'utilisation restrictive de ce modèle. En effet, les approches utilisées dans la déduction de la distribution de Wishart se basent surtout sur les aspects physiques des phénomènes de rétro-diffusion et de propagation des ondes radars réfléchies. Ainsi, des hypothèses ont été faites et des simplifications ont été introduites à travers les différentes étapes de ces approches. Ceci rend leurs résultats moins convaincants.

Du point de vue pratique, les difficultés rencontrées dans la mise en œuvre du modèle de Wishart, dû aux nombres et natures des paramètres à estimer, ont contribué dans la non-utilisation de ce modèle dans la pratique. La matrice de cohérence est obtenue par une transformation linéaire de la matrice de covariance polarimétrique. Cette dernière est estimée à partir de la matrice de rétro-diffusion de la cible. Pour chaque pixel de l'image POLSAR, le modèle de Wishart ainsi obtenu nécessite le calcul de la matrice de cohérence, sa matrice inverse, et son déterminant (expression (72)). En outre, pour une estimation précise et non singulière de la matrice de cohérence de la classe, un nombre suffisant et représentatif des échantillons d'apprentissage présélectionnés est requis. Ainsi, l'application du modèle devient une tâche très fastidieuse, coûteuse, et non pratique.

Mis à part toutes ces remarques sur le modèle de Wishart, selon les différents travaux publiés sur ce sujet [12]-[16], [68]-[71], on peut dire que la distribution de Wishart a été utilisée pour élaborer un classifieur de maximum de vraisemblances d'images POLSAR multivues plutôt que pour l'obtention d'un modèle statistique convenable pour l'analyse et la classification des images SAR d'une manière générale.

3. 2. Modélisation par la Distribution K

L'origine de l'utilisation de la distribution K pour la modélisation des échos radars revient à l'équipe britannique du radar royal "*British Team of the Royal Radar*" [60], [47]. En introduisant l'idée que plusieurs points de dispersion dans une cellule de résolution

fluctuent, la distribution K avec deux paramètres (dans sa forme originale) est obtenue, dans laquelle la distribution de Rayleigh est incluse comme un cas particulier [60]. Elle a été appliquée d'abord à la modélisation des échos radars d'un terrain naturel [55]. D'un point de vue mathématique, la distribution K est rigoureusement formulée, mais peut également être justifiée en termes de processus physiques de dispersion [11]. Elle se réduit à une distribution de Rayleigh dans des circonstances appropriées [56], [57]. Elle est particulièrement utile pour décrire des cibles spatialement étendue [59], [55], [9], [19], [47], [61] et récemment généralisée pour décrire des images POLSAR [58], [17], [62].

Basé sur une représentation mathématique rigoureuse de la distribution K, nous avons présenté dans notre publication [11] (voir l'annexe A pour la version intégrale de la publication) deux approches différentes pour la caractérisation statistique et la modélisation des images SAR. La première est une approche analytique, tandis que la deuxième est inspirée des techniques de traitement du signal. En effet, les modèles basés sur la distribution K développés dans la littérature sont donnés d'une façon clairsemée et ont besoin des éclaircissements sur la manière dont ils décrivent effectivement les images SAR aussi bien que la manière de les utiliser dans la pratique. Ceci est l'objectif de notre publication [11], en donnant ainsi une présentation unifiée sur l'obtention de la distribution K en tant que modèle statistique pour les images SAR et montrant, pour des fines pratiques, que les modèles dérivés à partir de la distribution K ont besoin d'estimer un seul paramètre à la fois.

Tout d'abord, la caractérisation statistique et la modélisation des images SAR monovues, basées sur la distribution K, sont décrites. Ensuite, la caractérisation statistique et la modélisation des images SAR multivues sont dérivées de celles des images SAR monovues. Pour les images POLSAR, la distribution K généralisée est donnée. Inspiré du principe du modèle de mélange de distributions, la deuxième approche est basée sur la transformation et la convolution multiplicative de Mellin pour la modélisation des images SAR. Un formalisme original pour la modélisation des images SAR, nous a permis d'approcher la caractérisation du processus de filtrage des images SAR en termes des statistiques du bruit speckle. Sachant que le modèle est multiplicatif, pour les images SAR et le speckle, et connaissant la distribution du speckle, cette approche nous a permis de réduire les calculs analytiques relatifs à la mise en œuvre de la distribution K en utilisant les transformées directe et inverse de Mellin, et la convolution multiplicative de Mellin. L'approche a l'avantage par rapport à la première en menant à une détermination plus facile des différents modèles pour les images SAR et le speckle.

Contrairement à la dérivation originale de la distribution K [60], [47], qui est en fonction de deux paramètres, les modèles décrits dans [11] dépendent d'un seul paramètre dont l'estimation est améliorée en utilisant la méthode d'échantillonnage bootstrap couplée à la technique de Monte Carlo. L'estimation de ce paramètre peut être effectuée en utilisant

une des trois méthodes d'estimation statistique suivantes : La méthode des moments, en utilisant les moments échantillonnés d'ordre deux et quatre [63], [64], la méthode de Ragahavan, basée sur la moyenne arithmétique et géométrique des échantillons [65], et la méthode de maximum de vraisemblances, recommandée quand la forme de la distribution à estimer est bien connue [18]. Le choix de l'estimateur dépend d'une part de la forme exacte de la distribution K , si elle est très proche de la distribution de Rayleigh ou bien si elle est légèrement non-Rayleigh, et d'autre part de la taille de l'ensemble des échantillons indépendants utilisé dans l'estimation [11].

Pour confirmer le choix judicieux de la distribution K , nous avons effectué une étude comparative entre la distribution K et d'autres distributions. Nous avons choisi les trois distributions les plus utilisées dans la littérature, qui sont les distributions de Gamma, Weibull, et Lognormal. Une représentation adéquate de ces distributions dans le système de Pearson nous a permis de donner des interprétations physiques de ces modèles. Le but est de comparer ces distributions et la distribution K afin de déterminer celle qui est la plus appropriée pour la caractérisation statistique des images SAR.

4. Conclusion

Dans ce chapitre, nous avons abordé la modélisation des images de télédétection civile, optiques et radars, nécessaire pour l'analyse et classification de ces images.

Le modèle Gaussien pour les images optiques est alors proposé et l'estimation de ses paramètres statistiques est discutée. Une méthodologie pour la sélection des échantillons plus ou moins statistiquement indépendants et représentatifs est donnée. Pour les images multispectrales de dimension faible, la méthode d'estimation des statistiques d'ordre un et deux au sens de maximum de vraisemblances est détaillée. Dans le cas des images multispectrales de dimension élevée, de nouveaux estimateurs de la matrice de covariance ont été proposés dans la littérature. Après un état de l'art sur les différents estimateurs proposés, nous avons exposé les estimateurs les plus récents, en faisant ressortir leurs avantages et leurs inconvénients afin de donner une vision plus claire sur la tendance actuelle de la recherche pour résoudre ce problème.

Pour la modélisation d'images radars, nous avons insisté sur deux modèles probabilistes à savoir la distribution de Wishart et la distribution K . Nous avons donné la manière dont la distribution de Wishart a été déduite comme modèle pour les cibles étendues dans les images POLSAR multivues, puis nous avons justifié pourquoi cette distribution ne peut pas être considérée comme modèle pour les images SAR. Contrairement à la distribution de Wishart, la distribution K a été largement utilisée dans la communauté. Basé sur une représentation mathématique rigoureuse de la distribution K , deux approches différentes

pour la caractérisation statistique et la modélisation des images SAR ont été proposées dans notre publication [11]. Les modèles décrits dans [11] dépendent d'un seul paramètre dont l'estimation est améliorée en utilisant la méthode d'échantillonnage bootstrap couplée à la technique de Monte Carlo. D'autres modèles probabilistes utilisés dans la littérature pour la modélisation d'images SAR ne sont pas donnés dans ce chapitre, mais sont comparés à la distribution K [11].

Références

- [1] J. A. Richards, *Remote Sensing Digital Image Analysis*, 2nd Edition, Springer-Verlag, Berlin, 1993.
- [2] J. P. Hoffbeck and D. Landgrebe, *Classification of High Dimensional Multispectral Data*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-EE95-14, May 1995.
- [3] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*, McGraw-Hill, 1978.
- [4] S. Chitroub, *Implémentation d'un Processus d'Analyse et de Classification d'Images Satellitaires. Application à la Caractérisation des Etats de Surface*, Thèse de Magister, Institut d'Electronique, U.S.T.H.B, Alger, 03/94-M/EL, May 1994.
- [5] G. B. Goldstein, "False-alarm regulation in Lognormal and Weibull clutter", *IEEE Transaction on Aerospace and Electronic systems*, Vol. AES-9, pp. 84-92, January 1973.
- [6] R. T. Frankot and R. Challappa, "Lognormal random-field models and their applications to radar images synthesis", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. GE-25, No. 2, pp. 195-207, March 1991.
- [7] A. L. Maffett and C. C. Wackerman, "The modified beta density function as a model for SAR clutter statistics", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 29, No. 2, pp. 277-283, 1991.
- [8] H. Laur, *Analyse d'Images Radars en Télédétection: Discriminateurs Radiométriques et Texturaux*, Thèse de Doctorat, Université de Paul Sabatier, Toulouse, France, 1989.
- [9] Z. Belhadj, *Apport de la Polarimetrie Multifréquence pour la Classification en Télédétection Radar*, Thèse de Doctorat, Ecole Doctorale Sciences pour l'Ingénieur de Nantes, France, 1995.
- [10] C. Oliver and S. Quegan S., *Understanding Synthetic Aperture Radar Images*, Artech-House, London, 1998.
- [11] S. Chitroub, A. Houacine, and B. Sansal, "Statistical characterisation and modelling of SAR images", *Signal Processing*, Vol. 82, No. 1, pp. 69-92, February 2002.
- [12] S. R. Cloude and E. Pottier, "An Entropy based classification scheme for land applications of polarimetric SAR", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 35, No. 1, pp. 68-78, January 1997.
- [13] J.S. Lee, M.R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on the complex Wishart distribution", *International Journal of Remote Sensing*, Vol. 15, No. 11, pp. 2299-2311. 1994.
- [14] J.S. Lee, M.R. Grunes, T.L. Ainsworth, L. Du, D.L. Schuler, and S.R. Cloude, "Unsupervised classification using polarimetric decomposition and complex Wishart distribution", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, N°. 5, pp. 2249-2258, September 1999.
- [15] S. R. Cloude and E. Pottier, "A Review of target decomposition theorems in radar polarimetry", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 34, N°. 2, pp. 498-518, September 1995.
- [16] E. Pottier and J. S. Lee, "Application of the "H/A" polarimetric decomposition theorem for unsupervised classification of fully polarimetric SAR data based on the Wishart distribution",

- IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, N°. 5, pp. 2249-2258, September 1999.
- [17] S. R. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, H. A. Zebker, and T. Le Toan, "K distribution and multi-frequency polarimetric terrain radar clutter", *Journal of Electromagnetic Waves and Applications*, Vol. 5, No. 1, pp. 1-15, 1991.
 - [18] I. R. Joughin, D. P. Percival, and P. Dale, "Maximum likelihood estimation of K-distribution parameters for SAR data", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 31, No. 5, pp. 989-999, 1993.
 - [19] R. Joughin, D. P. Percival, and P. Dale, "Probability density functions for multilook polarimetric signatures", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 32, No. 3, pp. 562-574, 1994.
 - [20] S. Quegan and I. Rhodes, "Statistical models for polarimetric data: consequences, testing and validity", *International Journal of Remote Sensing*, Vol. 32, 1994.
 - [21] J. M. Nicolas, M. Sigelle, C. Thuillier, and F. Tupin, "Images de radar à ouverture synthétique: transformée de Mellin et multiresolution", In *Proc. 16th Conference GRETSI'97*, Grenoble, France, Sept. 15-19, pp. 797-800, 1997.
 - [22] S. Chitroub, R. Hamiti, S. Allal, A. Houacine, and B. Sansal, "Supervised fusion-classification of multifrequency polarimetric SAR images using K distribution and theory of evidence", In *Proc. IEEE 1999 International Geoscience and Remote Sensing Symposium (IGARSS'99)*, Hamburg, Germany, 28 June-2 July, pp. 2001-2004, 1999.
 - [23] P. M. Mather, *Computer Processing of Remotely-Sensed Images*, John Wiley & Sons, 1987.
 - [24] A. J. Richard and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall International, New-Jersey, 1990.
 - [25] J.H. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, Vol. 84, pp. 165-175, 1989.
 - [26] J.P. Hoffbeck and D.A. Landgrebe, "Covariance matrix estimation and classification with limited training data", *IEEE Transaction of Pattern Analysis and Machine Intelligence*, Vol. 18, No. 7, pp. 763-767, 1996.
 - [27] C. Lee and D. Landgrebe, *Feature Extraction and Classification Algorithms for High Dimensional Data*, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, TR-EE 93-1, January 1993.
 - [28] C. Lee and D. Landgrebe, "Analyzing high-dimensional multispectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 4, pp. 792-800, July 1993.
 - [29] C. Lee and D.A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Transaction of Pattern Analysis and Machine Intelligence*, Vol. 15, No. 3, pp. 388-400, 1993.
 - [30] S. Tadjudin, *Classification of High Dimensional Data with Limited Training Samples*, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, 1998.
 - [31] S. Tadjudin and D. A. Landgrebe, "Covariance estimation with limited training samples", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 4, pp. 2113-2118, July 1999.
 - [32] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd Ed., New York, John Wiley & Sons, USA, 1984.
 - [33] S.P. Lin and M.D. Perlman, "A Monte Carlo comparison of four estimators of a covariance matrix", *Multivariate Analysis-VI : Proceedings of the Sixth International Symposium on Multivariate Analysis*, P.R. Krishnaiah, ed., Amsterdam: Elsevier Science Pub. Co., pp. 411-429, 1985.
 - [34] P.W. Wahl and R.A. Kronmall, "Discriminant functions when covariances are equal and sample sizes are moderate," *Biometrics*, Vol. 33, pp. 479-484, 1977.
 - [35] W. Rayens and T. Greene, "Covariance pooling and stabilization for classification", *Computational Statistics and Data Analysis*, Vol. 11, pp. 17-42, 1991.
 - [36] T. Greene and W. S. Rayens, "Partially pooled covariance matrix estimation in discriminant analysis", *Communications in Statistics – Theory and Methods*, Vol. 18, N°. 10, pp. 3679-3702, 1989.

-
- [37] K. K. Dey and C. Srinivasan, "Estimation of a covariance matrix under Stien's loss", *Annals of Statistics*, Vol. 13, N° 4, pp. 1581-1591, 1985.
 - [38] L. R. Haff, "Empirical Bayes estimation of the multivariate normal covariance matrix", *Annals of Statistics*, Vol. 8, N° 3, pp. 586-597, 1980.
 - [39] B. Efron and C. Morris, "Multivariate empirical Bayes and estimation of covariance matrices", *Annals of Statistics*, Vol. 4, N° 1, pp. 22-32, 1976.
 - [40] B. W. Flury, M. J. Schmid, A. Narayanan, "Error rates in quadratic discrimination with constraints on the covariance matrices", *Journal of Classification*, Vol. 11, pp. 101-120, 1994.
 - [41] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Ed., Academic Press, Boston, 1990.
 - [42] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd Ed., Johns Hopkins University Press, Baltimore, 1989, P. 51.
 - [43] F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications*, Artech House, 1990.
 - [44] J. C. Dainty, *Laser Speckle and Related Phenomena*, Springer-Verlag, New York, 1975.
 - [45] J. W. Goodman, "Some fundamental properties of speckle", *Journal of American Optical Society*, Vol. 66, pp. 1145-1149, 1976.
 - [46] L. Tsang, J. A. Kong, and R. T. Shin, *Theory of Microwave Remote Sensing*, Wiley & Sons, New York, 1985.
 - [47] C. J. Oliver, "The interpretation and simulation of clutter textures in coherent images", *Inverse Problem*, Vol. 2, pp. 481-518, 1986.
 - [48] M. C. Dobson, F. T. Ulaby, and L. E. Pierce, "Land-cover classification and estimation of terrain attributes using synthetic aperture radar", *Remote Sensing Environment*, Vol. 5, pp. 199-214, 1995.
 - [49] J. J. Van Zyl, H. A. Zebker, and C. Elachi, "Imaging radar polarization signatures: theory and observation", *Radio Science*, Vol. 22, No. 4, pp. 529-543, 1987.
 - [50] H. A. Zebker, J. J. Van Zyl, and D. N. Held, "Imaging radar polarimetry from wave synthesis", *Journal of Geophysical Research*, Vol. 92, No. 1, pp. 683-701, 1987.
 - [51] J. J. Van Zyl, "The effect of topography on radar scattering from vegetated areas", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 31, No. 1, pp. 153-160, 1993.
 - [52] J. A. Kong, A. A. Swartz and H. A. Yueh, "Identification of terrain cover using the optimum polarimetric classifier", *Journal of Electromagnetic Wave Application*, Vol. 30, No. 5, pp. 973-980, 1992.
 - [53] N. R. Goodman, "Statistical analysis based on a certain multiplicative complex Gaussian distribution (an introduction)", *Annals of Mathematical Statistics*, Vol. 34, pp. 152-177, 1963.
 - [54] M. Borgeaud, R. T. Shin, and J. A. Kong, "Theoretical models for polarimetric radar clutter", *Journal of Electromagnetic Waves and Applications*, Vol. 1, pp. 67-86, 1987.
 - [55] J. K. Jao, "Amplitude distribution of composite terrain radar clutter and the K-distribution", *IEEE Transactions on Antenna and Propagation*, Vol. 32, No. 10, pp. 1049-1062, 1984.
 - [56] E. Jakeman, "On the statistics of K-distribution noise", *Journal of Applied Physics*, Vol. 13, pp.31-48, 1980.
 - [57] E. Jakeman and R. J. A. Tough, "Generalized K distribution: A statistical model for weak scattering", *Journal of American Optical Society*, Vol. 4, No. 9, pp. 1764-1772, 1987.
 - [58] S. H. Yueh, J. A. Kong, J. K. Jao, R. T. Shin, and L. M. Novak, "K-distribution and polarimetric terrain clutter", *Journal of Electromagnetic Waves and Applications*, Vol. 3, No. 8, pp. 747-768, 1989.
 - [59] F. T. Ulaby, F. Kouyate, B. Brisco, and T. H. Lee, "Textural information in SAR images", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 24, pp. 235-245, 1986.
 - [60] E. Jakeman and P. N. Pusey, "A model for non Rayleigh sea echo", *IEEE Transactions on Antenna and Propagation*, Vol. 24, No. 6, pp. 806-814, 1976.
 - [61] A. Lopes, H. Laur, and E. Nezry, "Statistical distribution and texture in multilook and complex SAR images", In *Proceedings of IEEE 1990 International Geoscience and Remote Sensing Symposium (IGARSS'90)*, Washington, May 20-24, 1990.
-

-
- [62] L. M. Novack, M. C. Burl, and W. W. Irvig, "Optimal polarimetric processing for enhanced target detection", *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 5, pp. 234-244, 1993.
 - [63] C. Oliver, "Information from SAR images", *Journal of Applied Physics*, Vol. 24, pp. 1493-1514, 1991.
 - [64] S. H. Yueh, J. A. Kong, J. K. Jao, and R. T. Shin, "Statistical modeling for polarimetric remote sensing of earth terrain", in *Proceedings of IEEE 1990 International Geoscience and Remote Sensing Symposium (IGARSS'90)*, Washington, pp. 157-160, May 20-24, 1990.
 - [65] R. S. Raghavan, "A method for estimating parameters of K-distributed clutter", *IEEE Transactions on Aerospace Electronic Systems*, Vol. 27, No. 2, pp. 238-246, 1991.
 - [66] H. A. Zebker and J. J. Van Zyl, "Imaging radar polarimetry: a review", In *proceedings of the IEEE*, Vol. 79, No. 11, November 1991.
 - [67] S. Le Hégarat-Masclé, *Classification non supervisée d'images SAR polarimétriques*. Thèse présentée pour obtenir le grade de docteur de l'Ecole Nationale Supérieure des Télécommunications, ENST 96 E 024, septembre 1996.
 - [68] F. Sery, D. Ducrot-Gambart, A. Lopes, R. Fjørtoft, E. Cubero-Castan, P. Marthon, "Multisource classification of SAR images with the use of segmentation, polarimetry, texture and multitemporal data", in *Proc. SPIE, EUROPTO Series, Image and Signal Processing Conference*, Vol. 2955, Taormina, Italy, September 1996.
 - [69] J. S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interference SAR imagery", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 5, pp. 1017-1028, July 1994.
 - [70] W. M. Boerner *et al.*, "Polarimetry in radar remote sensing: Basic and applied concepts," in *The Manual of Remote Sensing*, 3rd ed. New York: Wiley, 1998, ch. 5.
 - [71] J. S. Lee, M. R. Grunes, and G. de Grandi, "Polarimetric SAR speckle filtering and its implication for classification", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 5, pp. 2363-2373, September 1999.
 - [72] J. W. Goodman, *Statistical Optics*. Wiley & Sons, New York, 1985.
 - [73] B. C. Kuo and D. Landgrebe, *Improved Statistics Estimation and Feature Extraction for Hyperspectral data Classification*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-ECE 01-6, December 2001.
 - [74] Q. Z. Jackson and D. Landgrebe, *Design of an Adaptive Classification Procedure for the Analysis of High-Dimensional Data with Limited Training Samples*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-ECE 01-5, December 2001.

Chapitre 3

Analyse Multidimensionnelle d'Images Optiques et Radars

1. Introduction

Les images multispectrales fournies par les satellites optiques de télédétection satisfont une grande variété d'applications [1]-[10]. Elles constituent la solution efficace et économique pour avoir l'information fiable et récente sur n'importe quelle zone d'intérêt. Cependant, la qualité de telles images dépend des conditions de travail du satellite. Elles peuvent subir des déformations et des perturbations dues aux effets atmosphériques, aux vibrations des capteurs du système d'acquisition, et autres. De plus, le rapprochement des parties du spectre électromagnétique correspondant aux capteurs du système d'acquisition du satellite se traduit par des images spectrales corrélées. Le traitement de toutes les images spectrales implique donc un certain degré de redondance. Si la taille de la scène observée et/ou le nombre d'images spectrales croît, l'espace mémoire pour le stockage des données devient important et le coût de traitement augmente. Donc il est nécessaire d'avoir une nouvelle représentation optimale des images multispectrales.

Le paragraphe 2 de ce chapitre est consacré à l'analyse multidimensionnelle d'images optiques. Les différentes méthodes proposées dans la littérature peuvent être divisées en deux catégories : Méthodes de sélection des caractéristiques discriminantes [9]-[22], en se basant sur le pouvoir discriminant entre les classes fourni par les images spectrales ; et méthodes d'extraction de caractéristiques via une transformation linéaire des images spectrales [9]-[10], [23]-[43]. Cette transformation linéaire peut être déduite par plusieurs approches. Ici, seules les méthodes d'extraction de caractéristiques basées sur l'Analyse en Composantes Principales (ACP) et ses variantes seront considérées. En suite, les approches neuronales de l'ACP seront exposées [44]-[49]. Le recours aux approches neuronales de l'ACP sera justifié. Les autres méthodes d'extraction de caractéristiques ne seront pas considérées mais le lecteur peut se référer aux références correspondantes citées dans la bibliographie.

Les capteurs radars présentent des avantages par rapport aux capteurs optiques [50]. Plusieurs systèmes de télédétection radar sont devenus maintenant opérationnels, tels que : la série de satellites ERS “*European Remote Sensing satellites*” ; RadarSAT ; et EnviSAT. La dernière décennie a vu le développement d’une nouvelle génération de radars multifréquences et multipolarisations tels que : AirSAR et SIR-C/X [51], [52]. L’apport d’informations fourni par ces nouveaux instruments, par rapport au cas de radars monofréquences et monopolarisations est très efficace pour une meilleure caractérisation des interactions de l’onde électromagnétique avec la surface imagée [50]-[53].

Cependant, l’analyse et l’interprétation des images radars sont plus compliquées que l’analyse et l’interprétation des images optiques. Les images radars possèdent des propriétés particulières qui demandent des techniques d’analyse et d’interprétation fondamentalement différentes de celles utilisées pour les images optiques. En tenant compte de ces particularités, nous avons proposé une nouvelle méthode pour l’analyse multidimensionnelle des images POLSAR [57], [58]. La méthode est basée sur une nouvelle forme de l’ACP qui fait appel à un théorème algébrique sur le calcul matriciel. Les problèmes qui se posent dans l’analyse des images radars seront exposés dans le paragraphe 3. Ensuite, un état de l’art sur les différentes approches proposées dans la littérature pour l’analyse des images POLSAR sera donné, suivi d’une synthèse générale de notre travail publié dans [58].

2. Analyse Multidimensionnelle d’Images Optiques

Afin d’avoir une représentation optimale de l’information portée par les images multispectrales fournies par les satellites dans un espace de dimension réduite et sans distorsion de cette information, plusieurs méthodes et approches ont été utilisées. Cependant, on peut grouper ces méthodes et approches en deux catégories selon la manière avec laquelle cette représentation optimale est obtenue : Méthodes de sélection des caractéristiques ; et méthodes d’extraction de caractéristiques.

La première catégorie de méthodes consiste à faire la sélection des caractéristiques en se basant sur le pouvoir discriminant entre les classes fourni par les images spectrales. Ces approches sont basées sur les mesures de séparabilité statistique entre les classes, en utilisant des distances statistiques [9]-[14], [17], [18], [20] et l’évaluation de l’erreur de classification [9], [13]-[16], [19]-[22]. Elles sont utilisées dans le but d’améliorer la classification des images de télédétection. Cependant, ces approches nécessitent l’estimation des matrices de covariance intra-classes et inter-classes à partir des échantillons d’apprentissage tirés des différentes classes. Les probabilités *a priori* des classes jouent un rôle important dans les performances de ces approches. Pour assurer des estimateurs exacts, il faut donc extraire un nombre suffisant d’échantillons représentatifs

des classes. Si le nombre de classes et/ou la dimension de l'image multispectrale est important, l'extraction des échantillons d'apprentissage devient une tâche fastidieuse et parfois impossible. D'autant plus que pour des classes de formes étroites, par exemple les routes et les fleuves, il est très difficile d'estimer leurs probabilités *a priori*. Par conséquent les résultats des classifications, en utilisant les caractéristiques sélectionnées par ces approches, ne sont pas précis pour ces types de classes. Pour ces raisons, ces approches ne sont pas très utilisées en pratique.

Dans la deuxième catégorie de méthodes, une manière d'améliorer la répartition des données est de les projeter sur un autre espace multidimensionnel de dimension faible, dans lequel la dimension intrinsèque des données originales est retenue. La dimension intrinsèque est définie comme étant le nombre de nouvelles caractéristiques nécessaire pour générer ou représenter l'ensemble global des données originales [59]. Ces caractéristiques saillantes peuvent être extraites via des transformations, telles que la transformée de Fourier, la transformée de cosinus discret [9], l'analyse discriminante linéaire [11], l'analyse canonique [10], [14], et l'ACP [9], [10]. Ces approches consistent à transformer les données d'un espace de représentation de dimension élevée en un autre espace de dimension inférieure où les données originales peuvent être reconstruites avec un minimum de déformations. Pour les transformées de Fourier et cosinus discret, le critère utilisé est déduit de l'aspect spatial (fréquentiel) des pixels de l'image. L'analyse discriminante linéaire et l'analyse canonique sont basées sur des critères dépendant de l'erreur de classification, et par conséquent du type de classifieur utilisé pour la classification des images. Le concept de leurs matrices de transformation linéaire fait appel aux matrices de covariance intra-classes et inter-classes et les probabilités *a priori* des classes. Ainsi, les performances de ces approches sont fortement dépendantes du processus d'extraction d'un nombre suffisant d'échantillons d'apprentissage représentatifs des classes et de l'évaluation précise de l'erreur de classification. Pratiquement, ceci est très difficile à assurer. D'autres méthodes, qui ne seront pas détaillées ici, sont dédiées à l'analyse multidimensionnelle des images multispectrales de dimension élevée [33], [34], [37], [43]. Ces méthodes utilisent aussi des matrices estimées des échantillons d'apprentissage, néanmoins leur concept est basé sur le fait que l'effet du problème d'insuffisance des échantillons d'apprentissage n'influe pas sur les performances de ces méthodes.

L'ACP présente l'avantage d'être simple dans sa mise en œuvre, par rapport aux méthodes précitées, et optimale dans le sens où elle compresse les données au sens des moindres carrées et réalise en même temps la décorrélation des images. Pour ces raisons, l'ACP est la méthode généralement la plus utilisée dans les applications de télédétection. Elle consiste à concentrer les variances des images spectrales, qui caractérisent l'information portée par l'image multispectrale, dans un petit ensemble de nouvelles images appelées images Composantes Principales (CP) dont la corrélation est nulle et dont le rapport signal sur bruit "*Signal to Noise Ratio (SNR)* en anglais" est largement amélioré

[10], [16], [36], [40]. Dans les deux sous-paragraphes suivants, nous allons donner l'approche statistique et les approches neuronales de l'ACP et ses variantes. Nous allons aussi justifier le pourquoi du recours aux approches neuronales et pourquoi elles sont avantageuses par rapport à l'approche statistique.

2. 1. Analyse en Composantes Principales : Approche Statistique

Avec l'ACP, nous voulons atteindre deux principaux objectifs : Le premier objectif est de concentrer l'information qui existe dans les différentes images spectrales, constituant l'image multispectrale fournie par le satellite, dans un nombre plus petit de nouvelles images CP améliorées, c-à-d. ayant des valeurs de SNR importantes. Le deuxième objectif est d'avoir des images CP décorréliées.

2. 1. 1. ACP standard appliquée en télédétection

Soient X le vecteur d'images spectrales, Y le vecteur d'images CP et A la matrice de la transformation linéaire qui nous permet d'obtenir Y à partir de X , nous avons alors :

$$Y = A.X \quad (1)$$

Par définition, la matrice de covariance du vecteur Y est donnée par :

$$S_Y = E(Y.Y^T) = E(A.X(A.X)^T) = E(A.X.X^T.A^T) = A.E(X.X^T).A^T = A.S_X.A^T \quad (2)$$

où S_X est la matrice de covariance du vecteur X qui est une matrice symétrique définie positive. Par conséquent, les valeurs propres de S_X sont toutes positives : $l_i > 0$, pour $i = 1, 2, \dots, p$, avec p est la dimension du vecteur X (nombre d'images spectrales). Soient $d_1, d_2, \dots, d_p$ les vecteurs propres normalisés associés aux valeurs propres l_i . En utilisant la notion de décomposition spectrale d'une matrice symétrique, nous allons montrer que S_Y est la forme diagonale de S_X . La décomposition spectrale de la matrice S_X est donnée par (pour la démonstration voir [10]) :

$$S_X = \sum_{i=1}^p l_i.d_i.d_i^T \quad (3)$$

$$\text{avec : } \begin{cases} d_i^T.d_j = 1 & \text{pour } i = j \\ d_i^T.d_j = 0 & \text{pour } i \neq j \end{cases} \quad (4)$$

Soit D la matrice orthogonale formée par les vecteurs propres normalisés $d_1, d_2, \dots, d_p$, correspondant aux valeurs propres l_i , nous avons alors :

$$S_X = D.L.D^T \quad (5)$$

où L est une matrice diagonale dont les éléments de la diagonale sont les valeurs propres l_i . Comme $D.D^T = D^T.D = I$, où I est la matrice identité, en posant $A = D^T$ nous obtenons :

$$S_Y = A.S_X.A^T = A.D.L.D^T.A^T = D^T.D.L.D^T.D = I.L.I = L \quad (6)$$

Ainsi, pour le choix $A = D^T$, la matrice de covariance des nouvelles images CP est la forme diagonale de S_X . La matrice de la transformation linéaire A est donc la transposée de la matrice des vecteurs propres normalisés de S_X . La $i^{ème}$ image CP est alors donnée par l'expression :

$$y_i = d_i^T.X \quad (7)$$

La variance de l'image CP y_i est telle que (pour la démonstration voir [10]) :

$$var(y_i) = d_i^T.S_X.d_i = l_i \quad (8)$$

Les variances des images CP sont alors égales aux valeurs propres correspondantes. Donc l'information portée par les images spectrales est conservée dans les nouvelles images CP. En fait, puisque la somme des variances des images spectrales, x_i , est égale la trace de la matrice S_X , nous pouvons donc facilement montrer que la somme des variances des images spectrales est égale à la somme des variances des images CP :

$$\sum_{i=1}^p var(x_i) = Tr(S_X) = Tr(D.L.D^T) = Tr(L.D.D^T) = Tr(L) = \sum_{i=1}^p l_i = \sum_{i=1}^p var(y_i) \quad (9)$$

Nous montrons maintenant que la covariance entre deux images CP est nulle. Pour deux vecteurs propres orthonormés, d_i et d_k , nous pouvons avoir les relations suivantes :

$$S_X.d_k = l_k.d_k \quad (10)$$

$$d_i^T.S_X.d_k = l_k.d_i^T.d_k = 0 \quad (11)$$

$$cov(y_i, y_k) = E((d_i^T.X)(d_k^T.X)^T) = E(d_i^T.X.X^T.d_k) = d_i^T.E(X.X^T).d_k = d_i^T.S_X.d_k \quad (12)$$

De l'expression (11) nous déduisons que la covariance entre deux images CP est nulle : $cov(y_i, y_k) = 0$. Ainsi, les images CP obtenues sont décorréées. Cependant, il faut noter que les vecteurs propres S_X sont orthogonaux si toutes les valeurs propres correspondantes sont distinctes. S'il y a certaines valeurs propres qui sont égales, (faute de quoi la matrice S_X est dite mal conditionnée pour des raisons pratiques), le choix des vecteurs propres correspondant n'est pas unique. Néanmoins, les vecteurs propres correspondant aux valeurs propres communes peuvent être choisis de façon à être orthogonaux [60].

Supposons que le bruit additif dans les images spectrales, issu principalement des dispositifs électroniques du système d'acquisition et des effets atmosphériques [9], [10],

soit non corrélé aux valeurs radiométriques des pixels. Supposons aussi que ce bruit est caractérisé par un vecteur moyenne nul et une matrice de covariance diagonale dont les éléments diagonaux sont égaux (variance unique égale à s_b^2). Autrement dit, soient B le vecteur du bruit additif et X_n le vecteur d'images spectrales non bruitées. Soient S_{X_n} et S_B leurs matrices de covariance, respectivement. Nous avons alors les relations suivantes :

$$X = X_n + B \quad (13)$$

$$S_X = S_{X_n} + S_B \quad (14)$$

$$E(B) = 0, \quad E(B.B^T) = s_b^2.I, \quad \text{et} \quad E(X_n.B^T) = 0 \quad (15)$$

A partir des équations (13), (14), et (15), nous allons montrer que les vecteurs propres qui forment la matrice de la transformation linéaire de l'ACP restent insensibles au bruit additif. Autrement dit, la transformation linéaire $Y = A.X$ n'a aucun effet sur les statistiques du bruit additif. D'après (14) et (15), nous avons les résultats suivants :

$$S_X.d_i = (S_{X_n} + S_B).d_i = (S_{X_n} + s_b^2.I).d_i \quad (16)$$

$$S_{X_n}.d_i = S_X.d_i - s_b^2.I.d_i = l_i.d_i - s_b^2.d_i = (l_i - s_b^2).d_i = l_i^n.d_i \quad (17)$$

où $l_i^n = (l_i - s_b^2)$ sont les valeurs propres de la matrice de covariance des images spectrales non bruitées. Nous déduisons de là que chaque image CP est affectée par le même bruit additif de départ de variance s_b^2 . Le rapport signal sur bruit (SNR) dans les images CP est alors égal à :

$$SNR = l_i/s_b^2 \quad (18)$$

Comme la matrice de covariance S_X est une matrice symétrique définie positive, donc ses valeurs propres sont telles que [60] : $l_1 > l_2 > \dots > l_p > 0$. Par conséquent, les images CP sont classées par variance décroissante et l'amélioration du SNR est alors effectuée dans les premières images CP, dont les valeurs propres correspondantes sont très grandes par rapport au reste des valeurs propres. Les dernières images CP, dont les valeurs propres correspondantes sont trop faibles, sont très bruitées. Nous pouvons dire que la proportion de l'information portée par les images spectrales et contenue dans la $k^{ème}$ image CP, notée R_k , est donnée par le rapport suivant :

$$R_k = (l_k) / \left(\sum_{i=1}^p l_i \right) \quad (19)$$

Par conséquent, les premières images CP représentent un pourcentage élevé de l'information existante dans les images spectrales. Ces images ont des valeurs de SNR élevées. De ce fait, la réduction de la dimensionalité avec amélioration du contenu et qualité des images est assurée en retenant seulement les premières images CP.

Nous présentons ici le résultat d'une ACP effectuée sur une image multispectrale (500x500 pixels) extraite d'une scène acquise par le satellite Landsat 7 en Février 2002 de résolution spatiale au sol 30x30 mètres (tous droits réservés au fournisseur *Eurimage*). La scène couvre une région située entre Novi Sad et Belgrade en Yougoslavie.

Les six images spectrales sont données dans les figures 1, 2, 3, 4, 5, et 6, respectivement. Les six images CP obtenues sont données dans les figures 7, 8, 9, 10, 11, et 12, respectivement. Le tableau 1 présente les coefficients de corrélation entre les six images spectrales. Le tableau 2 nous donne, pour chaque image CP, les coefficients du vecteur propre correspondant. Ces coefficients nous donnent une idée sur l'importance de chaque image spectrale dans la formation de l'image CP. Nous présentons dans le tableau 3 la proportion de l'information en pourcentage portée par chaque image CP par rapport à l'information globale existante dans les images spectrales (P). Nous remarquons que l'information globale existante dans les images spectrales est approximativement contenue dans les trois premières images CP (98.4%). Les valeurs du SNR dans les images spectrales et les images CP sont données dans les tableaux 4 et 5, respectivement.

La première image CP est une image de qualité très contrastée. Elle correspond pratiquement à la brillance globale de la scène observée et nous donne une idée générale sur sa topographie. La valeur du SNR dans cette image est largement supérieure par rapport à la meilleure valeur du SNR dans les images spectrales (à savoir la septième image spectrale). La deuxième image CP est moins contrastée et un peu bruitée. L'information portée par cette image, de variance égale à 282.98, exprime les écarts des différents types de la brillance globale de la scène, en particulier les éléments linéaires de la scène (fleuve, route, et les bords des champs cultivés) qui sont bien localisés dans cette image. Dans la troisième image CP, de variance égale 94.52, nous remarquons bien la prédominance du bruit par rapport à l'information. Les trois dernières images CP ne présentent que du bruit, où nous observons bien l'effet de lignage des capteurs du système d'acquisition dans la cinquième et sixième image CP. Ainsi, si nous retenons seulement les deux premières images CP, dont la variance et la valeur de SNR sont importantes et la proportion de l'information portée est 98.4%, nous pouvons dire que nous avons réalisé une réduction de la dimensionalité avec une amélioration dans le contenu et la qualité des images. Ces images CP retenues peuvent être utilisées pour d'autres fins, telles que l'interprétation de la scène imagée et la classification de la scène.

2. 1. 2. Variantes de l'ACP appliquées en télédétection

Dans les applications de télédétection et selon les objectifs à réaliser, plusieurs variantes de l'ACP ont été développées dans la littérature [10], [25]-[29], [31], [32], [35], [36], [38]-[42]. Mathématiquement, ces variantes sont justifiées, puisque l'existence de la matrice de la transformation linéaire est prouvée [10], [36], [38], [40], [41]. Pratiquement, ces

variantes ont montré leur efficacité pour réaliser les objectifs pour lesquels elles ont été développées. Nous donnons dans ce qui suit ces variantes de l'ACP. Le formalisme mathématique de chaque variante ainsi que son but d'application sont aussi abordés.

2. 1. 2. a- ACP standardisée :

Les résultats d'une ACP obtenus à partir de la décomposition spectrale de la matrice de covariance des images spectrales sont fortement dépendants des variances des différentes images spectrales. Une légère modification dans les valeurs des variances des images spectrales a pour effet de changer les proportions de l'information portée par chaque image CP ainsi que sa valeur du SNR. Par conséquent, l'interprétation des résultats obtenus est à revoir [10], [25], [31], [32], [38], [60], [61]. En fait, les images spectrales de variances importantes sont pondérées par des coefficients élevés (en valeurs absolues) dans la formation des premières images CP. Tandis que les images spectrales de variances relativement faibles participent à la formation des premières images CP avec de faibles coefficients, ainsi qu'à la formation des dernières images CP par des coefficients plus ou moins importants (voir les tableaux 1 et 2). Or, la variance statistique d'une image spectrale est liée aux réponses spectrales des objets terrestres dans la partie du spectre électromagnétique correspondante. Une variance importante ne traduit pas forcément l'abondance spatiale de la scène. De même, une variance relativement petite peut être très utile pour la cartographie des classes étroites, telles que les fleuves, les routes...etc. Autrement dit, il est parfois intéressant d'équilibrer entre les participations des différentes images spectrales dans la formation des premières images CP, afin de permettre aux images spectrales de variances relativement faibles d'avoir plus de chance d'introduire l'information qu'elles portent dans les premières images CP. Ceci est possible, en mettant toutes les images spectrales possédant la même moyenne statistique et la même variance statistique. Nous effectuons alors une ACP des images spectrales centrées et réduites. Les images CP obtenues sont appelées images CP standardisées.

Soit Z le vecteur des images spectrales centrées réduites :

$$Z = (S^{1/2})^{-1} X \quad (20)$$

où $S^{1/2}$ est une matrice diagonale dont les éléments sont les racines carrées des écart-types des images spectrales constituant le vecteur X . La matrice de covariance du vecteur Z est alors égale :

$$S_Z = E(Z Z^T) = E\left\{ (S^{1/2})^{-1} X X^T (S^{1/2})^{-1} \right\} = (S^{1/2})^{-1} E(X X^T) (S^{1/2})^{-1} \quad (21)$$

$$\text{Sachant que : } ((S^{1/2})^{-1})^T = ((S^{1/2})^T)^{-1} \text{ et } (S^{1/2})^T = S^{1/2} \Rightarrow ((S^{1/2})^{-1})^T = (S^{1/2})^{-1}$$

Par conséquent, la matrice de covariance du vecteur Z est égale à la matrice de corrélation du vecteur X , notée C_X , autrement dit :

$$\mathbf{S}_z = (\mathbf{S}^{1/2})^{-1} \cdot \mathbf{S}_x (\mathbf{S}^{1/2})^{-1} = \mathbf{C}_x \quad (22)$$

Une ACP des images spectrales centrées et réduites revient donc à dériver la matrice de la transformation linéaire à partir de la matrice de corrélation des images spectrales originales. Les images CP standardisées ainsi obtenues ont donc une moyenne statistique nulle et une variance proche ou inférieure à 1. Ces images CP standardisées peuvent être très utiles pour l'analyse multi-temporelle de la scène observée [27]. Bien que l'application de l'ACP des images spectrales centrées et réduites ait prouvé son efficacité [25], [32], [38], il faut, cependant, interpréter les résultats obtenus avec précaution, puisque l'ACP est une méthode qui est fortement dépendante de la nature des données à analyser [60].

Nous présentons dans les figures 13, 14, 15, 16, 17, et 18 les six images CP standardisées obtenues à partir d'une ACP des images spectrales centrées et réduites de la région extraite d'une scène acquise par le satellite Landsat 7 en Février 2002 (figures 1, 2, 3, 4, 5, et 6). La matrice de transformation linéaire de l'ACP des images centrées et réduites est donnée dans le tableau 6. Dans le tableau 7, nous présentons la proportion de l'information en pourcentage portée par chaque image CP standardisée par rapport à l'information globale existante dans les images spectrales (P). Ces proportions sont différentes de celles des images CP de l'ACP standard (voir tableau 3). Nous remarquons pour la première image CP standardisée que les coefficients correspondants aux images spectrales sont presque de même grandeur, alors que dans la première image CP de l'ACP standard (tableau 1), les coefficients des images spectrales ayant de variances importantes sont plus forts que ceux des images spectrales de variances moyennes ou faibles. Cette différence entre les deux images est visible à l'œil nu puisque l'image de l'ACP des images centrées et réduites apparaît un peu plus contrastée. Concernant la deuxième image CP, la comparaison entre les deux images des deux ACP montre que l'aspect visuel de l'image de l'ACP des images centrées et réduites est un peu amélioré par rapport à celui de l'image de l'ACP standard. Ceci peut être justifié par le fait que les coefficients correspondant aux images spectrales sont différents et que la valeur de SNR de l'image de l'ACP des images centrées et réduites, tableau 8, est plus élevée par rapport à celle de l'image de l'ACP standard, tableau 5. Pour le reste des images CP, il est inutile de faire la comparaison puisque le bruit est dominant dans les images dans les deux ACP.

2. 1. 2. b- ACP sélective :

Avec l'ACP standard, il arrive parfois que l'information contenue dans les dernières images CP, extrêmement bruitées, soit de grand intérêt. Cette information, caractérisée par de petites variances, peut nous donner une idée plus claire sur les structures fines de la scène imagée telles que : les structures géologiques, les réseaux routiers, les fleuves, détection de changement...etc. Elle représente le contraste spectral existant entre les différentes images spectrales. Ainsi, au lieu de chercher des images CP ayant des variances importantes, il est d'intérêt de chercher l'information inhérente à chaque image spectrale.

L'ACP sélective fait appel au même logiciel développé pour l'ACP standard. La différence entre les deux est que la première utilise toutes les images spectrales comme entrée, alors que la deuxième utilise seulement un sous-groupe d'images spectrales sélectionnées. La sélection des images spectrales dépend de l'objectif visé. Cet objectif est réalisé en exploitant le contraste spectral existant entre les images spectrales. Le critère de sélection est basé sur la corrélation : une corrélation forte entre les images spectrales implique que le contraste spectral est faible et une corrélation moyenne ou faible implique que le contraste spectral est important. Cependant, il faut noter que les régions du spectre électromagnétique correspondant aux images spectrales influent non seulement sur le degré de corrélation entre ces images, (le contraste spectral), mais aussi sur la cause de ce contraste.

Si le but est de réduire la dimensionnalité avec un minimum d'informations contenues dans les dernières images CP bruitées, alors le sous-groupe d'images spectrales fortement corrélées est utilisé comme entrée de l'ACP. L'information commune portée par ces images sélectionnées sera contenue dans la première image CP et le reste des images CP ne contient que du bruit. Si le but est la caractérisation du contraste spectral entre les images spectrales, de sorte que l'information inhérente à chaque image spectrale puisse être obtenue, les sous-groupes choisis doivent contenir des paires d'images spectrales de corrélation moyenne ou faible. Dans ce cas là, seulement la deuxième image CP est d'intérêt puisqu'elle contient le contraste entre les deux images spectrales. Cette image est moins bruitée puisque la corrélation entre les deux images d'entrée est moyenne ou faible. La première image CP contient l'information commune entre les deux images spectrales.

Les figures 19, 20, et 21, présentent les résultats d'une ACP sélective obtenus en utilisant les bandes spectrales 1, 2 et 3 fortement corrélées de la région située entre Novi Sad et Belgrade (voir le tableau 1). Le choix de ces images vise à cartographier l'information commune dans la première image CP. Cette image, nette et contrastée, offre un aperçu général de toute la topographie de la scène imagée. Sa proportion d'informations est de l'ordre de 97% par rapport à l'information portée par les bandes spectrales 1, 2, et 3. Les classes de superficies étendues sont bien retrouvées dans cette image. En revanche, les deux images CP suivantes sont dominées par le bruit et l'information est difficile à extraire de ces images. Nous présentons dans les figures 22 et 23, les résultats d'une ACP sélective où cette fois, les deux images d'entrée sélectionnées ont une corrélation faible (la plus faible entre les bandes spectrales originales, voir tableau 1). Ces images sont les bandes spectrales 2 et 4. Nous remarquons que la deuxième image CP est légèrement bruitée et reflète les différences existantes entre les deux bandes spectrales 2 et 4. Les structures linéaires de la scène, notamment les bords des champs cultivés et les routes, sont bien localisées dans cette image. Nous pouvons dire que visuellement, il est plus facile d'interpréter les résultats d'une ACP sélective que ceux d'une ACP standard lorsqu'il s'agit d'une scène dont les types de couverture terrestre sont de natures très variées.

2. 1. 2. c- ACP à bruit ajusté :

Le bruit dans les images multispectrales fournies par les satellites de la série Landsat et la série SPOT est supposé additif, Gaussien, non-corrélé et ayant les mêmes statistiques [10]. En appliquant l'ACP standard, ou bien ces deux variantes citées précédemment, le bruit est relégué aux dernières images CP [10], [16], [36], [38]. Cependant, dans les images multispectrales fournies par la nouvelle génération de satellites de télédétection civile, tel que le système AVIRIS, la variance de bruit dans certaines images spectrales est comparable aux variances propres des autres images spectrales. Il y a donc des statistiques différentes à travers les images spectrales. L'ACP peut alors forcer la présence du bruit dans les premières images CP et donne des images CP mêlées en terme de qualité d'image.

Des exemples de ce comportement ont été discutés dans [28], [29], et [39]. Une méthode dérivée de l'ACP appelée transformation de maximum de fractions de bruit "*Maximum Noise Fraction (MNF) transform*" s'avère utile pour analyser un tel type d'images multispectrales [28], bien que son utilisation exige l'estimation de la matrice de covariance du bruit. Les images CP de la transformation MNF sont classées par ordre de qualité, c.-à-d. par les valeurs des SNR correspondantes. Cette transformation a été reformulée plus tard comme étant une ACP à bruit ajusté "*Noise-Adjusted Principal Component (NAPC) transform*" [29]. La matrice de la transformation linéaire de NAPC est la solution d'un problème de valeurs propres généralisé qui nécessite la diagonalisation simultanée des deux matrices de covariance des images spectrales et du bruit additif [35].

En exploitant les résultats théoriques publiés dans [35], une généralisation de l'ACP adaptée à la nature d'images spectrales affectées par un bruit additif variable a été proposée [41]. Le concept de la matrice de transformation linéaire est déduite à partir de la matrice de covariance des images non bruitées et celle du bruit additif. La matrice de covariance du bruit est obtenue à partir des images spectrales filtrées dans le domaine spatial [23], [24]. Cette matrice se transforme en une matrice identité après transformation. Par conséquent, les images CP obtenues sont affectées par un même bruit statistiquement non corrélé de variance égale à 1. La matrice de covariance des images spectrales se transforme en même temps en une matrice diagonale dont les éléments diagonaux sont classés dans l'ordre décroissant. Les images CP sont alors non corrélées et le SNR est amélioré dans les premières images CP de variances importantes. Nous présentons dans ce qui suit l'analyse théorique de cette méthode ainsi que sa mise en œuvre.

- Analyse théorique :

Soit X le vecteur constitué par les p images spectrales, X_n le vecteur constitué par les p images spectrales non bruitées et B le bruit additif de statistiques différentes à travers les images spectrales. Soient S_X , S_{X_n} , et S_B les matrices de covariance des vecteurs X , X_n et B , respectivement. Nous avons alors la relation suivante entre les trois matrices de covariance :

$$\mathbf{S}_X = \mathbf{S}_{X_n} + \mathbf{S}_B \quad (23)$$

Soit Y le vecteur résultat de la transformation linéaire $Y = \mathbf{A}X$ où $\mathbf{A}$ est la matrice de la transformation linéaire recherchée. Soit $\mathbf{S}_Y$ la matrice de covariance du vecteur Y , nous avons alors :

$$\mathbf{S}_Y = \mathbf{A}^T \mathbf{S}_X \mathbf{A} = \mathbf{A}^T \mathbf{S}_{X_n} \mathbf{A} + \mathbf{A}^T \mathbf{S}_B \mathbf{A} = \mathbf{S}_{Y_n} + \mathbf{S}_{Y_B} \quad (24)$$

La détermination de la matrice $\mathbf{A}$ doit satisfaire aux deux critères suivants :

- (1) La matrice $\mathbf{S}_{Y_n}$ est une matrice diagonale dont les éléments de la diagonale sont classés dans l'ordre décroissant.
- (2) La matrice $\mathbf{S}_{Y_B}$ est une matrice identité.

Le premier critère peut être reformulée par une fonction homogène f de degré zéro en chaque vecteur ligne, a_i , de $\mathbf{A}$ telle que :

$$l_i = f(a_i) = (a_i^T \mathbf{S}_{X_n} a_i) / (a_i^T \mathbf{S}_B a_i) \text{ et } \partial f(a_i) / \partial a_i = 0 \quad (25)$$

En généralisant l'équation (25) pour le reste des vecteurs lignes de la matrice $\mathbf{A}$, nous obtenons l'équation suivante :

$$(\mathbf{S}_{X_n} - \mathbf{S}_B \mathbf{L}) \mathbf{A} = 0 \quad (26)$$

où $\mathbf{L}$ est la matrice diagonale dont les éléments de la diagonale sont les l_i trouvées. La détermination de la matrice $\mathbf{A}$, consiste alors à résoudre l'équation (26) en tenant compte de la contrainte de blanchiment de la matrice de covariance du bruit, $\mathbf{S}_B$, c.-à-d. que l'équation suivante doit être vérifiée :

$$\mathbf{A}^T \mathbf{S}_B \mathbf{A} = \mathbf{I} \quad (27)$$

Plusieurs algorithmes ont été proposés dans la littérature pour la mise en œuvre de l'équation (26) en tenant compte de la contrainte (27) [28], [29], [23], [24]. Mais aucune de ces références n'a donné la justification théorique de l'algorithme proposé, par conséquent des difficultés ont été rencontrées dans l'implémentation et l'application efficace de ces algorithmes. Roger [35] a montré que le fondement théorique de ces algorithmes proposés est le théorème de diagonalisation simultanée de deux matrices par une seule matrice orthogonale [12], mais aucun résultat pratique n'a été donné. Dans ce qui suit, nous allons décrire la démarche théorique qui nous permet d'obtenir la matrice $\mathbf{A}$, via la résolution de l'équation (26) en tenant compte de la contrainte (27). La démarche nous permettra aussi de déduire les étapes pratiques nécessaires pour la mise en œuvre de la méthode et l'extraction des images CP à bruit ajusté.

- Extraction des images CP à bruit ajusté :

Pour commencer, les deux hypothèses suivantes sont supposées vraies :

- (1) La matrice S_B est supposée réelle, symétrique, et définie positive.
- (2) Au moins l'une des deux matrices S_{X_n} et S_B est supposée non singulière.

Suivant l'hypothèse (1), S_B admet la décomposition de *Cholesky* telle que :

$$S_B = H^T.H \quad (28)$$

où H^T est la matrice triangulaire inférieure et H est la matrice triangulaire supérieure. Soit le vecteur v_i donné par la relation suivante :

$$v_i = H.a_i \quad (29)$$

En remplaçant (28) et (29) dans (25) nous obtenons :

$$l_i = \left(v_i^T (H^{-1})^T . S_{X_n} (H^{-1}) v_i \right) / (v_i^T . v_i) \quad (30)$$

Pour le choix $v_i^T . v_i = 1$, nous obtenons :

$$\left((H^{-1})^T . S_{X_n} (H^{-1}) \right) v_i = l_i . v_i \quad (31)$$

Par conséquent, v_i est le vecteur propre de la matrice $\left((H^{-1})^T . S_{X_n} (H^{-1}) \right)$ associé à la valeur propre l_i . Comme les deux matrices S_B et $\left((H^{-1})^T . S_{X_n} (H^{-1}) \right)$ sont deux matrices symétriques réelles définies positives, elles possèdent les mêmes valeurs propres distinctes positives, l_i , telles que : $l_1 > l_2 > \dots > l_p > 0$, et les vecteurs propres associés, v_i , sont alors orthonormés.

Soit V la matrice orthogonale formée par les vecteurs propres v_i et L la matrice diagonale des valeurs propres correspondante. Si nous remplaçons dans l'expression (24) les matrices A et A^T par les matrices $H^{-1}.V$ et $V^T(H^{-1})^T$, respectivement, nous obtenons les deux relations suivantes :

$$S_{Y_n} = A^T . S_{X_n} . A = V^T (H^{-1})^T . S_{X_n} . H^{-1} . V = V^T . L . V \quad (32)$$

$$S_{Y_b} = A^T . S_B . A = V^T (H^{-1})^T . S_B . H^{-1} . V = V^T (H^T)^{-1} . H^T . H . H^{-1} . V = V^T . V = I \quad (33)$$

D'après les expressions (32) et (33), la matrice de la transformation linéaire A recherchée, qui est la solution de l'équation (26) en tenant compte de la contrainte (27), est alors la matrice $H^{-1}.V$. Or cette matrice est une matrice orthogonale, par conséquent nous

avons : $(H^{-1}.V)^T = (H^{-1}.V)^{-1}$. La transposée de la matrice A permet alors de diagonaliser simultanément les deux matrices symétriques S_{X_n} et S_B .

L'extraction des images CP à bruit ajusté consiste à projeter les images spectrales dans la base orthonormée $A^T = (H^{-1}.V)^T$. Les nouvelles images CP ainsi obtenues sont alors non corrélées et de variances égales aux valeurs propres distinctes, I_i , de la matrice symétrique $((H^{-1})^T.S_{X_n}.(H^{-1}))$ (équations (31) et (32)). La matrice de covariance du bruit additif, affectant les images CP, est une matrice identité (équation (33)). Par conséquent, les images CP sont affectées par le même bruit de variance égale à 1. Puisque les valeurs propres de la matrice $((H^{-1})^T.S_{X_n}.(H^{-1}))$ sont classées dans l'ordre décroissant, l'amélioration du SNR est alors effectuée dans les premières images CP de variances importantes (valeurs propres importantes). Les dernières images CP sont dominées par le bruit et le SNR correspondant est très faible.

- Implémentation :

Selon les expressions (28), (29), (31), (32) et (33), l'implémentation de l'ACP à bruit ajusté peut être formulée par les étapes pratiques suivantes :

- (1) Calculer la décomposition de Cholesky de la matrice S_B (expression (28)).
- (2) Calculer le produit matriciel $((H^{-1})^T.S_{X_n}.(H^{-1}))$.
- (3) Calculer les matrices des vecteurs propres orthonormés, V , et des valeurs propres correspondantes, L , (expression (31)).
- (4) Calculer la matrice de la transformation linéaire de l'ACP à bruit ajusté, A^T , telle que : $A^T = V^T.(H^{-1})^T$.
- (5) Les images CP (vecteur Y) sont obtenues telles que : $Y = A^T.X$.

- Résultats pratiques :

En raison de la non-disponibilité d'images multispectrales du système AVIRIS, dans lesquelles la variance du bruit dans certaines images spectrales est comparable aux variances propres des autres images spectrales, nos expériences sont faites sur l'image multispectrale du satellite Landsat 7 (figures 1, 2, 3, 4, 5, et 6) dont nous avons rajouté un bruit additif de statistiques différentes. Les statistiques du bruit ajouté à travers les images spectrales sont résumées dans le tableau 9. Nous donnons dans la figure 24 l'image bruitée de la septième bande spectrale qui représente l'image spectrale originale ayant la valeur de SNR la plus élevée (voir tableau 4). La première image CP à bruit ajusté est donnée dans la figure 25. Afin d'effectuer une étude comparative, l'ACP standard a été appliquée sur les images spectrales bruitées. La première image CP obtenue est donnée dans la figure 26. La première image CP de l'ACP à bruit ajusté est très contrastée. Le bruit de départ dans les

images spectrales est largement réduit. La première image CP de l'ACP standard est aussi une image contrastée, mais en comparaison avec l'image CP de l'ACP à bruit ajusté elle est un peu bruitée. Ceci peut être justifié par le fait que la matrice de transformation linéaire de l'ACP standard n'a aucun effet sur la matrice de covariance du bruit. En revanche, la matrice de la transformation linéaire de l'ACP à bruit ajusté permet de rendre la matrice de covariance du bruit égale à la matrice identité. Les mêmes remarques peuvent être formulées pour les deuxièmes et les troisièmes images CP des deux ACP. Le reste des images CP est dominé par le bruit et l'information portée est négligeable. Ces remarques sont confirmées par les valeurs des SNR dans les images spectrales bruitées et les images CP des deux ACP données dans les tableaux 10, 11 et 12, respectivement.

2. 1. 2. d- ACP à ajustement automatique de bruit :

Une autre variante de l'ACP convenable pour l'analyse des images spectrales fournies par le système AVIRIS a été proposée dans [39]. Elle a la même forme que la transformation en CP standardisées. La différence entre les deux transformations est que les images spectrales ne sont pas cette fois ci centrées et réduites, mais elles sont pondérées par une matrice diagonale. Cette matrice est le produit de l'inverse de la matrice diagonale des écarts type des images spectrales par la matrice diagonale des coefficients de corrélation multiples [60] obtenus en considérant la régression linéaire multiple de chaque image spectrale sur le reste des images spectrales. Cette nouvelle forme de l'ACP ajuste d'une manière automatique le bruit existant dans chaque image spectrale. L'estimation de la matrice de covariance du bruit additif est effectuée à partir des images spectrales elles-mêmes via l'inverse de leur matrice de covariance. La matrice de transformation linéaire est alors égale à la matrice des vecteurs propres normalisés de la matrice de covariance des images spectrales pondérées. Nous exposons dans ce qui suit cette variante de l'ACP avec quelques commentaires sur les résultats pratiques publiés [39].

Soit à prédire l'image spectrale x_i à partir du reste des $(p-1)$ images spectrales, où p est la dimension de l'image multispectrale, en utilisant la prédiction linéaire, PL , telle que :

$$PL(x_i) = b_0 + b_{1,x_1} + \dots + b_{i-1,x_{i-1}} + b_{i+1,x_{i+1}} + \dots + b_{p,x_p} \quad (34)$$

L'erreur au sens des moindres carrés, notée e , d'une telle prédiction est donnée par :

$$e = E((x_i - PL(x_i))^2) \quad (35)$$

Cette manière de prédiction est considérée comme étant une régression linéaire multiple de l'image spectrale x_i sur le reste des $(p-1)$ images spectrales. Le degré de corrélation entre l'image spectrale x_i et le reste des $(p-1)$ images spectrales est exprimé par le coefficient de corrélation multiple donné par l'expression suivante :

$$0 \leq r_{i(p-1)}^2 = \frac{\mathbf{s}_{(p-i)i}^T \cdot \mathbf{S}_{(p-i)(p-i)}^{-1} \cdot \mathbf{s}_{(p-i)i}}{\mathbf{s}_{ii}} \leq 1 \quad (36)$$

$$\text{avec } \mathbf{s}_{(p-i)i} = (\mathbf{s}_{i1}, \mathbf{s}_{i2}, \dots, \mathbf{s}_{i(i-1)}, \mathbf{s}_{i(i+1)}, \dots, \mathbf{s}_{ip})^T \text{ et } \mathbf{S}_{(p-i)(p-i)} = \begin{pmatrix} \mathbf{s}_{ii} & \mathbf{s}_{(p-i)i}^T \\ \mathbf{s}_{(p-i)i} & \mathbf{s}_{ii} \end{pmatrix}$$

où $\mathbf{S}_{(p-i)(p-i)}$ est de rang plein, si ce n'est pas le cas ceci implique que parmi les $(p-1)$ images spectrales restantes, une image est une combinaison des autres. L'erreur au sens des moindres carrées de l'expression (35) peut être reformulée à partir de (36), telle que :

$$e = \mathbf{s}_{ii}(1 - r_{i(p-1)}^2) \quad (37)$$

Cette erreur peut être interprétée comme étant la variance résiduelle non prédite de l'image spectrale x_i . Or, nous ne pouvons pas prédire le bruit additif dans l'image spectrale x_i à partir du reste des images spectrales. Par conséquent, nous pouvons dire que la variance résiduelle caractérisée par l'erreur donnée dans (37) représente la variance du bruit dans l'image spectrale x_i . Ainsi, en généralisant les expressions (34), (36), et (37) pour toutes les images spectrales, la matrice de transformation linéaire de cette nouvelle variante de l'ACP peut être déduite de la manière suivante.

Soit U_s matrice diagonale dont les éléments de la diagonale sont les inverses des variances du bruit dans les images spectrales estimées à partir de l'erreur donnée en (37). Sous l'hypothèse que le bruit additif est décorrélié avec les contenus des images spectrales, alors la transformation linéaire de type :

$$X' = U_s \cdot X \quad (38)$$

est utilisée, où X est le vecteur des images spectrales originales, il permet de rendre la matrice diagonale du bruit, U_s , égale à une matrice identité. En appliquant ensuite une ACP standard sur le nouveau vecteur d'images spectrales, X' , nous obtenons le vecteur des images CP, Y , tel que :

$$Y = \mathbf{A} \cdot X' \quad (39)$$

où $\mathbf{A}$ est la matrice des vecteurs propres normalisés de la matrice de covariance du vecteur X' . Donc les images CP ainsi obtenues sont ordonnées en terme de valeurs de SNR. Autrement dit, la première image CP a la valeur de SNR la plus élevée, la deuxième image CP a la deuxième valeur de SNR la plus élevée, et ainsi de suite. De ce fait, le bruit additif de départ, dont les statistiques ne sont pas les mêmes à travers les images spectrales originales, est automatiquement ajusté après les deux transformations linéaires successives (38) et (39). Pour cette raison, cette nouvelle variante de l'ACP est appelée : ACP à ajustement automatique du bruit. Sa matrice de transformation linéaire, $\mathbf{A}'$, est telle que :

$$\mathbf{A}' = \mathbf{A} U_s \quad (40)$$

En utilisant des données spectrales réelles du système AVIRIS, une étude comparative entre l'ACP standard, la transformation en CP standardisées, et l'ACP à ajustement automatique du bruit a été effectuée dans [39]. Le résultat de l'étude a montré que l'ACP à ajustement automatique du bruit est la meilleure en termes de suppression du bruit et ordonnancement des images CP selon leurs valeurs de SNR. Cependant, en terme de compression de données (c.-à-d. retenir la dimension intrinsèque des données la plus réduite), la transformation en CP standardisées est la meilleure, suivie par l'ACP standard, puis l'ACP à ajustement automatique du bruit. Ceci a été justifié par le fait que la transformation en CP standardisées et l'ACP donnent des images CP mêlées en terme de qualité d'image (valeurs de SNR correspondantes). Par conséquent, des pertes d'informations ont lieu dans les images CP fortement bruitées.

2. 2. Analyse en Composantes Principales : Approches Neuronales

L'ACP standard et ses variantes sont basées sur la décomposition spectrale de la matrice de covariance des images spectrales d'entrée et celle du bruit additif pour les variantes qui font appel à cette matrice dans leur concept. Ceci consiste en premier lieu à estimer la matrice de covariance des images d'entrée, et celle du bruit pour certaines variantes, puis le calcul de leurs vecteurs propres et valeurs propres afin de former la matrice de transformation linéaire.

Cependant, dans la pratique, cette manière de mise en œuvre de l'ACP et de ses variantes souffre de la difficulté d'une évaluation précise des vecteurs propres qui forment la matrice de transformation linéaire. En particulier, lorsque les matrices de covariance des images d'entrée et du bruit sont mal conditionnées, pour des raisons pratiques, ou lorsque leur évaluation n'est pas précise. L'estimation des matrices de covariance n'est pas une tâche facile, en particulier lorsque la taille de l'image et/ ou le nombre d'images spectrales est important, où nous observons une accumulation des erreurs du calcul. Ces déficiences rendent l'ACP et ses variantes inefficaces pour des applications en télédétection. Ainsi, pour une application efficace de l'ACP et ses variantes, une méthode qui détermine avec précision la matrice de transformation linéaire, sans devoir estimer les matrices de covariance et calculer leurs décompositions spectrales, est nécessaire. Ceci peut être réalisé en utilisant les réseaux de neurones qui présentent le caractère d'auto-organisation dans l'apprentissage pour s'adapter à la nature des données.

Dans ce qui suit, après un état de l'art sur les modèles de réseaux de neurones de l'ACP développés dans les communautés de réseaux de neurones et de traitement du signal, les deux approches neuronales de l'ACP des images de télédétection que nous avons proposé dans [48] et [49] sont récapitulées dans les deux prochains sous-paragraphe.

2. 2. 1. Etat de l'art

Récemment, dans les communautés des réseaux de neurones et de traitement du signal, un certain nombre de modèles des réseaux de neurones pour l'ACP a été proposé [63], [64], [65]. Dans la plupart des cas, il a été montré que ces modèles ont de meilleures performances et surmontent les inconvénients inhérents aux approches classiques de l'ACP [66]. Cependant, à l'exception du travail développé par Lee et Landgrebe [67] pour la sélection des caractéristiques des données du système AVIRIS, aucune application de réseau neurones pour l'extraction de caractéristiques et réduction de dimension des images multispectrales n'a été rapportée dans la littérature.

La relation entre l'ACP et les réseaux de neurones a commencé quand il a été conjecturé que les neurones s'organisent individuellement pour identifier les caractéristiques spécifiques de l'environnement, et que cette propriété d'auto-organisation est liée aux critères statistiques, tel que celui de l'ACP. La plupart des travaux effectués dans ce contexte a été inspirée des systèmes biologiques tout en développant ce qu'on appelle l'apprentissage non-supervisé. Les règles d'apprentissage non-supervisé sont prévues pour apprendre à partir des observations (les données d'entrée). Le but est de découvrir les caractéristiques significatives des données d'entrée. Un des pionniers dans l'apprentissage non-supervisé était le psychologue Donald Hebb qui a formulé la règle Hebbienne d'apprentissage [68]. S'il y a deux neurones simultanément activés de chaque côté d'une connexion alors le poids de cette connexion est forcé. Pour bien illustrer la règle Hebbienne d'apprentissage, considérons un neurone simple de la figure 27, ayant comme entrée le vecteur $X = (x_1, x_2, \dots, x_p)^T$ et génère comme sortie la réponse y . Les connections sont orientées de l'entrée vers la sortie du neurone.

L'activation est calculée comme étant le produit inhérent du vecteur d'entrée X par le vecteur des poids synaptiques $W = (w_1, w_2, \dots, w_p)^T$, tel que :

$$u = \sum_{i=1}^p w_i \cdot x_i \quad (41)$$

La fonction d'activation, $y = f(u)$, calcule la sortie du neurone en fonction du produit inhérent. Hebb [68] a proposé une règle simple pour ajuster les poids synaptiques durant l'apprentissage du réseau : Quand deux neurones sont simultanément excités, la connexion entre eux est forcée. La fonction d'activation pourrait en principe être n'importe quelle fonction croissante du produit inhérent. Ainsi, selon le principe de la règle Hebbienne d'apprentissage, le changement dans les poids synaptiques dans un neurone simple est tel que :

$$Dw_i(k) = b_i y(k) x_i(k) \quad (42)$$

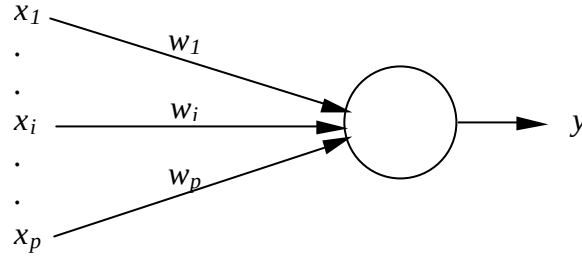


Figure 27. Neurone simple de p entrées et une sortie.

où k est l'itération et $\mathbf{b}$ est le paramètre d'apprentissage. La mise à jour des poids synaptiques prend alors la forme :

$$w_i(k+1) = w_i(k) + \mathbf{b} \cdot y(k) \cdot x_i(k) \quad (43)$$

Malheureusement, cette forme mène à une croissance exponentielle de w_i quand le nombre d'échantillons d'apprentissage augmente. Ceci a été résolu plus tard par Erkki Oja [69] en incorporant un terme de normalisation :

$$w_i(k+1) = \frac{w_i(k) + \mathbf{b} \cdot y(k) \cdot x_i(k)}{\left(\sum_i (w_i(k) + \mathbf{b} \cdot y(k) \cdot x_i(k))^2 \right)^{1/2}} \quad (44)$$

Pour de petites valeurs de paramètre d'apprentissage $\mathbf{b}$, l'équation (44) prend la forme approximative suivante :

$$w_i(k+1) = w_i(k) + \mathbf{b} \cdot y(k) (x_i(k) - y(k) \cdot w_i(k)) \quad (45)$$

L'équation (45) est appelée la règle d'apprentissage d'Oja et avec une petite modification, elle s'appelle aussi la règle Hebbienne d'apprentissage généralisée [49], [62]. Cette simple règle s'avère être étroitement liée à l'ACP quand les neurones sont représentés par un modèle linéaire. Oja [69] a prouvé qu'une version normalisée de la règle Hebbienne appliquée sur une simple unité linéaire converge vers le vecteur propre principal de la matrice de covariance des données d'entrée.

La normalisation proposée par Oja est non seulement essentielle pour la stabilité, mais constitue également la caractéristique commune dans plusieurs algorithmes neuronaux de l'ACP. Le rapport entre la règle Hebbienne normalisée et l'ACP, peut être développé et servir comme une idée de base pour plusieurs algorithmes neuronaux de l'ACP. Puisque la règle d'Oja seule pourrait seulement extraire une seule CP, les travaux développés ont étendu cette approche pour extraire plusieurs CP à partir d'un réseau de plusieurs neurones linéaires [70]-[76].

Le modèle d'extracteur adaptatif des CP "*Adaptive Principal-component Extractor (APEX)*" proposé dans [64], diffère de la plupart des autres modèles en utilisant ce qu'on

appelle les poids latéraux du réseaux qui servent à décorréler la $m^{ième}$ CP au cours de l'extraction avec les $(m-1)$ CP déjà extraites. Le modèle est basé sur le calcul récursif au sens des moindres carrés "*recursive least squares (RLS) computation*" pour extraire la $m^{ième}$ CP sachant que les $(m-1)$ CP ont été déjà extraites. La motivation derrière cette approche est la nécessité d'extraire les CP des données d'entrée dans les situations où le nombre nécessaire des CP à extraire est *a priori* inconnu. Parmi les modèles mentionnés ci-dessus, le modèle proposé dans [73] ne diffère pas de l'APEX à cet égard et possède presque la même architecture. Les deux modèles, cependant, diffèrent significativement en terme de rapidité de convergence due à l'utilisation de l'algorithme RLS par le modèle APEX. Ce modèle est également semblable au modèle proposé dans [65], qui utilise aussi l'algorithme RLS. Ce modèle n'a pas les connections latérales entre les neurones de sortie, mais plutôt utilise d'une manière explicite la transformation de déflation pour l'extraction des CP. Son avantage principal est que le paramètre d'apprentissage de l'algorithme de mise à jour n'est pas fixe et dépend des caractéristiques statistiques des données d'entrée. De ce fait, le modèle n'a pas le problème de précision engendré par la rapidité de convergence [69] et détermine d'une manière récursive la variance associée à chaque CP. Ce modèle a été utilisé [77] pour extraire les CP des petites fenêtres dans chaque bande d'une image multispectrale. Mais l'extraction des CP de l'image multispectrale entière n'a pas été considérée.

En résumé, les algorithmes neuronaux de l'ACP développés dans la littérature peuvent être divisés en deux catégories : Algorithmes de réestimation et algorithmes de décorrélation [62]. Dans les algorithmes de réestimation, le réseau de neurones a seulement les coefficients synaptiques connectant les neurones d'entrée aux neurones de sortie. La mise à jour de ces coefficients est basée sur la règle Hebbienne d'apprentissage. Avant d'extraire une nouvelle CP, on soustrait les premières CP déjà extraites de l'entrée avant que les données ne soient impliquées dans l'apprentissage. En revanche, dans les algorithmes de décorrélation, le réseau de neurones a les coefficients synaptiques et les coefficients latéraux. La mise à jour des coefficients synaptiques se fait selon la règle Hebbienne d'apprentissage, tandis que la mise à jour des coefficients latéraux est selon la règle anti-Hebbienne d'apprentissage. Les CP ainsi extraites sont décorrélées par annulation des coefficients latéraux durant l'apprentissage.

2. 2. 2. Approche basée sur l'algorithme d'apprentissage de décorrélation [48]

Nous avons proposé dans notre publication [48] (voir l'annexe A pour la version intégrale de la publication) un modèle de réseau neurones qui effectue l'ACP standard directement à partir des images spectrales sans aucune estimation préliminaire de la matrice de covariance des images spectrales ni le calcul de sa décomposition spectrale. Le modèle est basé sur l'algorithme d'apprentissage de type décorrélation. Son algorithme

d'apprentissage a été établi et implémenté de telle sorte que l'exécution du logiciel développé soit totalement indépendante de l'utilisateur. L'extraction des images CP désirées est séquentielle et adaptative. La convergence globale du modèle est étudiée. L'application du modèle a été réalisée sur une image multispectrale fournie par le satellite Landsat-TM. La capacité de compression de l'ACP neuronal proposé est prouvée en reconstituant les images spectrales avec un minimum d'erreurs au sens des moindres carrées. La précision des résultats obtenus est satisfaisante. L'étude comparative a prouvé que l'approche neuronale de l'ACP est supérieure à l'approche statistique de l'ACP standard.

2. 2. 3. Approche basée sur l'algorithme d'apprentissage de réestimation [49]

Le travail présenté dans notre publication [49] (voir l'annexe A pour la version intégrale de la publication) consiste à proposer un autre modèle de réseaux neurones qui permet d'effectuer l'ACP standard et ses variantes appliquées en télédétection. L'extraction des images CP est faite directement à partir des images spectrales sans aucun calcul non-neuronal additionnel et aucune estimation précédente d'une matrice quelconque. Le modèle est basé sur l'algorithme de réestimation et la règle Hebbienne d'apprentissage généralisée. En utilisant la technique de déflation pour calculer les vecteurs propres, le réseau extrait séquentiellement les images CP désirées en soustrayant les premières images CP déjà extraites de l'ensemble des images spectrales d'entrée avant que ces dernières soient impliquées dans l'apprentissage. L'optimisation du modèle est basée sur le critère de minimum d'erreurs moyennes carrées dans la reconstruction des images d'entrée à partir des images CP retenues. Son algorithme d'apprentissage est établi de telle sorte que l'exécution du logiciel développé soit totalement indépendante de l'utilisateur. Sa convenance pour les variantes de l'ACP est considérée. Sa convergence est étudiée et est acceptable. L'application du modèle a été réalisée sur une image multispectrale fournie par le satellite Landsat-TM. La capacité de compression du modèle proposé est prouvée via un processus de reconstruction des images spectrales avec un minimum de pertes d'information au sens des moindres carrées. La précision des résultats obtenus est satisfaisante. Une étude comparative entre le modèle neuronal proposé et l'approche statistique de l'ACP standard est également effectuée.

3. Analyse Multidimensionnelle d'Images SAR

Nous exposons dans ce qui suit les problèmes qui se posent dans l'analyse des images SAR. Ensuite, un état de l'art sur les différentes approches proposées dans la littérature pour l'analyse des images POLSAR est donné, suivie d'une synthèse générale de notre travail publié dans [58] (voir l'annexe A pour la version intégrale de cette publication).

3. 1. Problèmes posés en Analyse d'Images SAR

Les images SAR sont dégradées par le speckle. Le speckle est essentiellement dû à la cohérence des ondes radar et à la rugosité du terrain. C'est le résultat de l'interférence constructive et destructive entre les signaux réfléchis par les cibles ponctuelles. Il a les caractéristiques d'un bruit multiplicatif non-Gaussien [78]. Il apparaît comme un bruit granulaire qui dégrade l'aspect visuel de l'image et rend son analyse et son interprétation difficiles. Il présente des défis à la plupart des modèles pour le traitement et compréhension des images SAR. La réduction du speckle est donc nécessaire avant tout traitement efficace des images SAR.

Pour les images POLSAR, en plus du problème du speckle, l'information relative à la scène observée est dispersée dans plusieurs images polarimétriques. Ces images sont corrélées ce qui pose le problème de décorrélation de ces images afin de réduire la redondance d'information. L'autre problème posé avec les images POLSAR est le nombre d'images à traiter. Pour réduire l'espace mémoire et le temps de calcul, il est plus judicieux d'avoir un nombre plus faible de nouvelles images qui maximisent l'information portée par les images POLSAR. Ainsi, une représentation optimale de dimension faible (dans laquelle l'information est concentrée, les images sont décorréliées et l'effet du speckle est réduit) devient une tâche essentielle avant toute autre tâche de traitement.

3. 2. Etat de l'Art

L'ACP et ses variantes ne peuvent en aucun cas, être appliqués telles quelles pour l'analyse des images POLSAR [79]. Le modèle du bruit additif pour les images optiques n'est plus valable pour les images POLSAR affectées par le speckle.

Basé sur le principe que le speckle dans les images POLSAR est un bruit non-stationnaire, une méthode basée sur une transformation linéaire non-orthogonale appelée ACP généralisée a été proposée [80]. La méthode est une transformation en CP à bruit ajusté (NAPC) [29] adaptée aux images POLSAR [80] en supposant un modèle additif pour le speckle. Cependant, aucune analyse théorique justifiant la méthode ou une procédure d'implémentation claire n'ont été données. En conséquence, des ambiguïtés et des difficultés ont été rencontrées dans l'implémentation et l'application de la méthode. Selon les commentaires donnés dans [79], la méthode semble donner des résultats dégénérés dans le contexte du modèle multiplicatif, et il n'y a eu aucune évaluation détaillée de sa valeur pour l'analyse des images POLSAR.

Une autre méthode souvent utilisée dans les applications de télédétection radar consiste à appliquer la méthode standard de l'ACP dans le domaine logarithmique des images POLSAR. En prenant la forme logarithmique des images POLSAR, le speckle se

transforme en un bruit additif et par conséquent l'ACP standard peut être appliquée. Cependant, comme nous l'avons justifié dans notre publication [58], l'application de l'ACP dans le domaine logarithmique des images POLSAR ne donne pas de résultats significatifs et la perte d'information portée par les images SAR originales est clairement observée dans le processus de reconstitution des images POLSAR. Nous avons justifié ceci par le fait que l'opérateur logarithmique tend à diminuer, et donc à perdre l'information contenue dans les images, puisque l'opérateur logarithmique comprime la gamme dynamique des images. Les statistiques déterminées à partir de la forme logarithmique des images POLSAR sont donc complètement différentes de celles déterminées à partir des images POLSAR. Or dans l'ACP standard, la matrice de transformation linéaire est déduite de la décomposition spectrale de la statistique d'ordre deux des images d'entrée [10], [60].

3. 3. Méthode pour l'Analyse Multidimensionnelle d'Images POLSAR

La nouvelle méthode proposée dans notre publication [58] pour une représentation optimale des images POLSAR est basée sur une analyse théorique inspirée des aspects bien développés des théories des calculs matriciels. En convertissant le modèle multiplicatif du speckle en un modèle additif sans utiliser un opérateur de transformation, la méthode consiste à exploiter le théorème algébrique sur la diagonalisation simultanée de deux matrices carrées, symétriques, réelles et définies positives via une matrice carrée orthogonale. Les deux matrices à diagonaliser sont la matrice de covariance du bruit qui dépend des images et la matrice de covariance des images POLSAR. La matrice de covariance du bruit multiplicatif devient une matrice identité ce qui implique que le bruit est décorrélié des images et sa variance dans chaque nouvelle image est égale à 1. La matrice de covariance des images POLSAR se transforme en même temps en une matrice diagonale dont les éléments diagonaux sont ordonnés dans l'ordre décroissant. Ceci signifie que les nouvelles images obtenues sont décorréliées et classées par leurs variances (c.-à-d. par l'ordre de qualité). Du point de vue mathématique, la méthode est bien justifiée, ceci nous a facilité sa mise en œuvre. La procédure d'implémentation ainsi proposée est établie de telle sorte que l'application de la méthode soit pratique et simple.

La méthode a été testée et évaluée sur des images POLSAR réelles fournies par le système SIR-C/X. L'efficacité de la méthode du point de vue compression des données est prouvée en reconstituant les images POLSAR originales à partir d'un nombre restreint de nouvelles images avec une perte minimale d'information. Les résultats obtenus en termes de réduction du speckle et amélioration de la qualité des images sont satisfaisants. Les nouvelles images ainsi extraites sont appropriées pour une meilleure classification et interprétation de la scène imagée. De ce fait, la méthode est un outil puissant pour analyser des scènes complexes imagées par des capteurs SAR polarimétriques.

4. Conclusion

Les images optiques peuvent subir des déformations et des perturbations dues aux effets atmosphériques, aux vibrations des capteurs du système d'acquisition, et autres. Ceci se traduit par un bruit additif qui se rajoute au contenu réel du pixel. Le rapprochement des parties du spectre électromagnétique correspondant aux capteurs du système d'acquisition du satellite pose le problème de redondance d'information. Si la taille de la scène imagée et/ ou le nombre d'images spectrales croît, l'espace mémoire pour le stockage des données devient important et le coût de traitement augmente. Donc, une représentation optimale de dimension faible (dans laquelle l'information est concentrée, les images sont décorréelées, et l'effet du bruit est réduit) devient une tâche essentielle. Après avoir donné l'état de l'art sur les différentes méthodes de sélection et d'extraction des caractéristiques pertinentes, les approches statistiques de l'ACP et ses variantes appliquées en télédétection sont détaillées. Par la suite les approches neuronales proposées pour l'ACP et ses variantes sont exposées. Le recours vers les approches neuronales de l'ACP est justifié tout en faisant ressortir les avantages de ces approches.

En plus des problèmes de corrélation entre les images et la réduction de la dimensionalité, les images radars sont dégradées par le speckle qui a les caractéristiques d'un bruit multiplicatif non-Gaussien. La réduction du speckle est donc nécessaire avant tout traitement efficace des images SAR. Après avoir exposé les problèmes qui se posent dans l'analyse des images SAR, un état de l'art sur les différentes approches proposées dans la littérature pour l'analyse des images POLSAR est donné. Ensuite une nouvelle méthode proposée pour l'analyse multidimensionnelle des images POLSAR est exposée.

Références

- [1] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. Mc Graw-Hill, 1978.
- [2] M. F. Courel, *Etude de l'Evolution Recente des Milieux Sahéliens à Partir des Mesures Fournies par les Satellites*. Thèse de Doctorat d'Etat-Es-Lettres et Sciences Humaines, Université de Paris I, 1984.
- [3] T. M. Lillesand and R. W. Kiefer, *Remote Sensing and Image Interpretation*. John Wiley & Sons, 2nd Edition, 1987.
- [4] S. D. Drury, *Image Interpretation in Geology*. Allen & Unwin, London, 1987.
- [5] P. G. Pilon, P. J. Howarth, R. A. Bullock, and P. O. Adeniyi, "Problems of change detection in a semi-arid region of west Africa using Landsat-MSS data", *Presented at the Canadian Symposium on Remote Sensing*, University of Waterloo, Ontario, Jun 1987.
- [6] P. G. Pilon, P. J. Howarth, R. A. Bullock, and P. O. Adeniyi, "An Enhanced classification approach to change detection in semi-arid environments", *Photogrammetric Engineering & Remote Sensing*, Vol. 54, No. 12, pp. 1709-1716, December 1988.
- [7] R. Escadafal, *Caractérisation de la Surface des Sols Arides par Observation de Terrain et par Télédétection*. Ouvrage objet d'une Thèse de Doctorat d'Etat en Pédologie, Edition ORSTOM, Collection Etude et Thèses, Paris, 1989.

-
- [8] M. C. Girard & C. M. Girard, *Téledétection Appliquée Zones Tempérées et Intertropicales*. Edition Masson, Paris, 1989.
 - [9] J. A. Richards, *Remote Sensing Digital Image Analysis. An Introduction*, 2nd Edition, Springer-Verlag, Berlin, 1993.
 - [10] S. Chitroub, *Implémentation d'un processus d'analyse et de classification d'images satellitaires. Application à la caractérisation des états de surface*. Thèse de Magister, Institut d'Electronique, USTHB, Alger, 03/94-M/EL, May 1994.
 - [11] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.
 - [12] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd Edition, Academic Press, Boston, 1990.
 - [13] P. W. Mausel, W. J. Krambre, and K. Lee, "Optimum band selection for supervised classification of multispectral data", *Photogrammetric Engineering & Remote Sensing*, Vol. 57, No. 1, pp. 55-60, January 1990.
 - [14] B. Kim and D. A. Landgrebe, "Hierarchical classifier design in high-dimensional, numerous class cases", *IEEE Transactions on Geoscience & Remote Sensing*, Vol. 29, No. 4, pp. 518-528, July 1991.
 - [15] F. Csillag, L. Pásztor, and L. L. Beihl, "Spectral band selection for the characterization of salinity status of soils", *Remote Sensing of the Environment*, Vol. 43, 1993.
 - [16] S. Chitroub et A. Belhadj-Aissa, "Amélioration de la classification supervisée des données satellitaires par un choix optimal des caractéristiques", *14^{ème} colloque sur le Traitement du Signal et d'Images, GRETSI'93*, Juin-Les-Pins, du 13 au 15 septembre, 1993, Grenoble, France, pp. 173-176.
 - [17] J. C. Price, "Band selection procedure for multispectral scanners", *Applied Optics*, Vol. 33, No. 15, 1994.
 - [18] S. Chitroub, A. Belhadj-Aissa, et B. Sansal, "Analyse d'images satellitaires optiques par mesures de séparabilité statistiques", *5^{ème} Colloque Maghrébin sur les Modèles Numériques de l'Ingénieur, C2MNI5*, Ecole Mohammadia d'Ingénieur, Rabat, Maroc, du 21 au 23 Novembre, 1995, pp. 99-104.
 - [19] S. Chitroub, A. Belhadj-Aissa, et B. Sansal, "Evaluation de l'erreur de probabilité de la classification supervisée d'images satellitaires par mesures de séparabilité statistique", *Mediterranean Conference on Electronics and Automatic, MCEA'95*, September, 13-15 1995, Grenoble, France, pp. 35-40.
 - [20] S. Chitroub et B. Sansal, "Feature reduction in terms of higher statistical separability for enhancement of multispectral image classification", *International Conference on Electronics and Systems, ICECS'97*, December 15-18, 1997, Cairo, Egypt, pp. 1029-1034.
 - [21] M. Vélez-Reyes and L. O. Jiménez, "Subset selection analysis for the reduction of hyperspectral imagery", In *Proceedings of IEEE International Geosciences and Remote Sensing Symposium*, July 6-10, Seattle, WA, 1998.
 - [22] M. Vélez-Reyes and D. M. Linares, "Comparison of principal-component-based band selection methods for hyperspectral imagery", In *Proceedings of SPIE, Image and Signal Processing for Remote Sensing Conference*, September 21-25, 2002, Toulouse, France, Vol. 4541, pp. 361-369, 2002.
 - [23] P. Switzer and A. A. Green, "Min/Max autocorrelation factors for multivariate spatial imagery", *Technical Report*, Department of Statistics, Stanford University, California, USA, No. 6, April 1984.
 - [24] M. Berman, "The statistical properties of three noise removal procedures for multichannel remotely sensed data", *Consulting report*, NWS/85/31/MB9/1985, CSIRO division of Mathematics and Statistics, Australia, 1985.
 - [25] A. Singh and A. Harrison, "Standardized principal components", *International Journal of Remote Sensing*, Vol. 6, pp. 883-896, 1985.
 - [26] P. S. Chavez and A. Kwarteng, "Extracting spectral contrast in Landsat thematic mapper image data using selective principal component analysis", *Photometry Engineering and Remote Sensing*, Vol. 55, pp. 339-348, 1987.
-

-
- [27] T. Fung and E. LeDrew, "Application of principal components analysis to change detection", *Photometry Engineering and Remote Sensing*, Vol. 53, pp. 1649-1658, 1987.
 - [28] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal", *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 26, pp. 65-74, 1988.
 - [29] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal components transform", *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 28, pp. 295-304, 1990.
 - [30] S. C. Ahearns and C. Wee, "Data space volumes and classification optimization of SPOT and Landsat-TM data", *Photogrammetric Engineering & Remote Sensing*, Vol. 57, No. 1, pp. 61-65, January 1991.
 - [31] J. R. Eastman and M. Fulk, "Long sequence times series evaluation using standardized principal components", *Photogrammetric Engineering & Remote Sensing*, Vol. 59, No. 6, pp. 991-996, June 1993.
 - [32] L. Eklundh and A. Singh, "A comparative analysis of standardized and unstandardized principal components analysis in remote sensing", *International Journal of Remote Sensing*, Vol. 14, pp. 1359-1370, 1993.
 - [33] C. Lee and D.A. Landgrebe, "Feature extraction based on decision boundaries," *IEEE Transaction of Pattern Analysis and Machine Intelligence*, Vol. 15, No. 3, pp. 388-400, 1993.
 - [34] C. Lee and D. Landgrebe, *Feature Extraction and Classification Algorithms for High Dimensional Data*, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, TR-EE 93-1, January 1993.
 - [35] R. E. Roger, "A fast way to compute the noise-adjusted principal components transform matrix", *IEEE Transaction on Geoscience and Remote Sensing*, Vol. 32, pp. 1194-1196, 1994.
 - [36] S. Chitroub et A. Belhadj-Aissa, "Application de méthodes statistiques descriptives pour l'établissement de nouvelles représentations d'images satellitaires", *Journées de Statistiques Appliquées, JSA'94*, du 16 au 18 avril, 1994, Alger, Algérie, pp. 144-149.
 - [37] L. O. Jiménez and D. A. Landgrebe, *High Dimensional Feature Reduction via Projection Pursuit*, Ph. D. Thesis, TR-ECE, School of Electrical and Computer Engineering, Purdue University, West Lafayette, April 1995.
 - [38] S. Chitroub, A. Belhadj-Aissa, and B. Sansal, "Analysis of Landsat-TM images data using standardized and selective principal components transforms", In *Proceedings of Remote Sensing In Action (RSS'95)*, Southampton, England, 11-13 September 1995, pp. 1282-1289.
 - [39] R. E. Roger, "Principal components transform with simple, automatic noise adjustment", *International Journal of Remote Sensing*, Vol. 17, pp. 2719-2727, 1996.
 - [40] S. Chitroub et B. Sansal, "Principes des méthodes statistiques d'analyse de données images multidimensionnelles", *Conférence Maghrébine sur l'Automatique, l'Electrotechnique et l'Electronique Industriel, COMAEI'96*, du 3 au 5 Decembre, 1996, Tlemcen, Algérie, pp. 122-127.
 - [41] S. Chitroub et B. Sansal, "Analyse en composantes principales à bruit ajusté. Application compression et amélioration des images optiques de télédétection", *Journées de Statistiques Appliquées, JSA'97*, 30 novembre, 1-2 décembre 1997, Alger, Algérie, pp. 78-83.
 - [42] P. A. Siljeströme, A. Moreno, K. Vikgren, and L. M. Cacerres, "The application of selective principal components analysis (SPCA) to a Thematic Mapper (TM) image for the recognition of geomorphological features configuration", *International Journal of Remote Sensing*, Vol. 18, pp. 3843-3852, 1997.
 - [43] B. C. Kuo and D. Landgrebe, *Improved Statistics Estimation and Feature Extraction for Hyperspectral data Classification*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-ECE 01-6, December 2001.
 - [44] S. Chitroub, A. Houacine, A. El-Aziz Khadraoui, and B. Sansal, "Approches neuronales pour l'extraction des composantes principales d'images multispectrales de télédétection", *17ème Colloque sur le Traitement du Signal et d'Image, GRETSI'99*, Vanne, France, 13-17 septembre 1999, pp. 155-158.
-

- [45] S. Chitroub, A. Houacine, K. Ichir, N. Kamel, and B. Sansal, "Amélioration et compression d'image multispectrale en utilisant les réseaux de neurones", *Conférence sur la Vision Artificielle (CVA 2000)*, du 18 au 20 novembre 2000, Tizi-Ouzou, Algérie, pp. 103-108.
- [46] S. Chitroub, A. Houacine, N. Kamel, K. Ichir, and B. Sansal, "Compression and improvement of remotely sensed images using a PCA-based neural network model", *International Conference on Communication, Computer and Power (ICCCP'01)*, 12-14 February 2001, Muscat, Oman, pp. 102-106.
- [47] S. Chitroub, A. Houacine, and B. Sansal, "Principal component analysis of multispectral images using neural network", In *Proceedings of ACS / IEEE International Conference on Computer Systems and Application (AICCSA 2001)*, 26-29 June 2001, Beirut, Lebanon, pp. 105-111.
- [48] S. Chitroub, A. Houacine, and B. Sansal, "Neuronal principal component analysis for an optimal representation of multispectral images", *Intelligent Data Analysis, International Journal*, Vol. 5, No. 5, pp. 385-403, 2001.
- [49] S. Chitroub, A. Houacine, and B. Sansal, "Neural network model for standard PCA and its variants. Applied in remote sensing study", *Submitted to International Journal of Remote Sensing*.
- [50] S. Le Hégarat-Masclé, *Classification non Supervisée d'Images SAR Polarimétriques*. Thèse présentée pour obtenir le grade de docteur de l'Ecole Nationale Supérieure des Télécommunications, ENST 96 E 024, septembre 1996.
- [51] E. R. Stofan, D. L. Evans, C. Schmullius, B. Holt, J. J. Plaut, J. Van Zyl, S. D. Wall, J. Way, "Overview of results of spaceborne imaging radar-C, X-band synthetic aperture radar", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 33, No. 4, pp. 817-828, 1995.
- [52] J. Rolando, L. Bryan, L. Huneycutt, W. Marian, "The SIR-C/X SAR synthetic aperture radar system", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 33, No. 4, pp. 829-839, 1995.
- [53] F. T. Ulaby and C. Elachi, *Radar Polarimetry for Geoscience Applications*, Artech House, 1990.
- [54] C. Elachi, *Spaceborne Radar Remote Sensing: Applications and Techniques*. IEEE Press, New York, 1998.
- [55] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Artech-House, London, 1998.
- [56] S. Chitroub, A. Houacine, and B. Sansal, "Statistical characterisation and modelling of SAR images", *Signal Processing*, Vol. 82, No. 1, pp. 69-92, February 2002.
- [57] S. Chitroub et B. Sansal, "Extraction d'information, amélioration du SNR et compression des données dans les images SAR multifréquences multipolarisées", *16ème Colloque sur le Traitement du Signal et d'Image, GRETSI'97*, Grenoble, France, 13-17 septembre 1997, pp. 471-474.
- [58] S. Chitroub, A. Houacine, and B. Sansal, "A New PCA-based method for data compression and enhancement of multi-frequency polarimetric SAR imagery", *To be appear in Intelligent Data Analysis, International Journal*, Vol. 6, No. 2, 2002.
- [59] K. W. Pettits, T. A. Bailey, A. K. Jain and R. C. Dubes, "An intrinsic dimensionality estimator from near-neighbor information", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol. 1, pp. 25-37, 1979.
- [60] A. J. Richard, and D. W. Wichern, *Applied Multivariate Statistical Analysis*. 2nd Edition, Prentice Hall International, New Jersey, 1988.
- [61] W. P. Loughlin, "Principal component analysis for alteration mapping", *Photogrammetric Engineering & Remote Sensing*, pp. 1165-1169, January 1991.
- [62] S. Haykin, *Neural Networks, a Comprehensive Foundation*, 2nd Edition, Prentice-Hall, 1999.
- [63] E. Oja, "Principal components, minor components and linear neural networks", *Neural Networks*, Vol. 5, pp. 927-935, 1992.
- [64] S. Y. Kung, K. I. Diamantaras, K. I., and J. S. Taur, "Adaptive principal component extraction (APEX) and applications", *IEEE Transaction on Signal Processing*, Vol. 42, No. 5, pp. 1202-1217, 1994.

-
- [65] S. Bannour, and M. R. Azimi-Sadjadi, "Principal component extraction using recursive least squares learning", *IEEE Transaction on Neural Networks*, Vol. 6, No. 2, pp. 457-469, 1995.
- [66] J. Mao, and J. A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection", *IEEE Transaction on Neural Networks*, Vol. 6, No. 2, pp. 296-317, 1995.
- [67] C. Lee, and D. Landgrebe, "Decision boundary feature selection for neural networks", In *Proceedings of IEEE International Conference on Systems, Man and Cybernetics*, pp. 1053-1057, 1992.
- [68] D. O. Hebb, *The Organization of Behavior*, Wiley, New York, 1949.
- [69] E. A. Oja, "A simplified neuron model as a principal components analyzer", *Journal of Mathematical Biology*, Vol. 15, pp. 267-273, 1982.
- [70] E. Oja, "Neural networks, principal components, and subspaces", *International Journal of Neural Systems*, Vol. 1, pp. 61-68, 1989.
- [71] T. D. Sanger, "Optimal unsupervised learning in a single-layer linear feedforward neural network", *Neural Networks*, Vol. 2, pp. 459-473, 1989.
- [72] P. Földiák, "Adaptive network for optimal linear feature extraction", In *Proceedings of International Joint Conference on Neural Networks (IJCNN)*, Washington, DC, Vol. 1, pp. 401-405, 1989.
- [73] J. Rubner and K. Schulten, "Development of feature detectors by self-organization", *Biological Cybernetic*, Vol. 62, No. 6, pp. 193-199, 1990.
- [74] P. Baldi, and K. Hornik, "Neural networks and principal component analysis: Learning from examples without local minima", *Neural Networks*, Vol. 2, pp. 53-58, 1989.
- [75] S. Berker, "Unsupervised learning procedures for neural networks", *International Journal of Neural Systems*, Vol. 2, pp. 17-33, 1991.
- [76] E. Oja, "Principal components, minor components and linear neural networks", *Neural Networks*, Vol. 5, pp. 927-935, 1992.
- [77] X. Miao, M. R. Azimi-Sadjadi, B. Tian, A. C. Dubey, and H. Witherspoon Ned, "Detection of mines and minelike targets using principal components and neural-network methods", *IEEE Transaction on Neural Networks*, Vol. 9, pp. 454-463, 1998.
- [78] J.W. Goodman, "Some fundamental properties of speckle", *Journal of Optical Society*, Vol. 66, pp. 1145-1149, 1976.
- [79] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech-House, London, 1998.
- [80] J. S. Lee and K. Hoppel, "Principal component transformation of multifrequency polarimetric SAR imagery", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 30, No. 4, pp. 686-696, 1994.

Chapitre 4

Classification Ponctuelle d'Images de Télédétection

1. Introduction

L'utilisateur peut sur une image fournie par un satellite de télédétection civile distinguer certaines régions. Cependant, cette distinction est subjective, et d'autant plus délicate, dans le cas des images bruitées par un bruit additif ou multiplicatif, nuit à la lisibilité des images. La complexité des informations contenues dans les images de télédétection rend leur analyse par un photo-interprète lourde et ardue, voire impossible.

La classification d'images de télédétection est une analyse quantitative qui consiste à regrouper, dans une même entité ou classe, tous les pixels ayant la même signature spectrale ou polarimétrique tout en respectant leur localisation spatiale. Le résultat est une image des régions homogènes représentant des classes distinctes correspondant aux différents thèmes au sol. Cependant, le problème posé par la classification des images est la manière la plus efficace qui nous permet de traduire l'information contenue dans les images en un nombre de classes principales et faire correspondre ces classes aux différents types d'occupation du sol présents sur le site imagé. Les méthodes de classification des images de télédétection peuvent être divisées en méthodes supervisées et méthodes non-supervisées [1]-[3]. Les premières reposent sur une information *a priori* sur les classes importantes (leur nombre, leur signification, leurs caractéristiques statistiques), issue de bases de données, ou acquise lors d'une étape d'apprentissage, par opposition aux secondes qui ne nécessitent pas (ou très peu) d'informations *a priori* sur les classes présentes dans l'image. Le présent chapitre est essentiellement consacré à l'étude et au développement de méthodes de classification supervisée. Nous ne considérons pas les méthodes de classification non supervisée pour des raisons qui seront détaillées plus tard.

Afin de pouvoir évaluer le résultat d'une classification, nous avons besoin de méthodes d'évaluation de l'erreur de classification. Une manière précise d'évaluer l'erreur de classification est de comparer l'image résultat avec la réalité du terrain. Cependant, dans la

pratique il est très difficile, voir impossible, d'obtenir cette réalité du terrain. Généralement, on évalue la qualité d'une image classifiée par une appréciation visuelle qualitative suivie d'une appréciation quantitative basée sur un critère mathématique. Nous introduisons dans ce chapitre les différents critères permettant l'évaluation statistique d'une classification ainsi que la méthode adoptée pour la définition d'une réalité terrain.

La nouvelle génération des systèmes d'acquisition pour l'observation de la terre, tels que “*The Moderate Resolution Imaging Spectrometer (MODIS)*” [1] et AVIRIS [4]-[8], fournissent des images multispectrales de dimension élevée. Bien que les méthodes de classification principalement développées pour les images multispectrales de dimension faible puissent être utilisées, cependant, quelques problèmes qui ne sont pas rencontrés dans le cas de la dimension faible ont été soulevés lors de l'application de ces méthodes sur les images multispectrales de dimension élevée. A travers les nouveaux classifieurs rapides développés récemment pour cette fin, nous adressons certains de ces problèmes. En particulier, nous étudions l'importance relative des statistiques du premier et second ordre dans la discrimination entre les classes. L'influence des exactitudes des estimateurs de ces statistiques dans la précision de la classification est également abordée.

2. Classification Non-Supervisée

Dans les méthodes de classification non supervisée (appelée aussi classification automatique) [10], la loi de la distribution des observations conditionnellement à celle des classes, ainsi que la loi *a priori* de l'image des classes, sont modélisées *a priori*, et le caractère non supervisé des méthodes porte sur l'estimation des caractéristiques statistiques des classes par des processus mathématiques de regroupement de données [1], [2], [9]-[12]. Nous exposons dans le sous-paragraphe suivant quelques approches de classification non supervisée qui ont été déjà développées [13]-[16] à partir desquelles nous donnons les avantages et les inconvénients des méthodes de classification non-supervisée.

2. 1. Approches de Classification Non-Supervisée

La classification non-supervisée est basée sur le principe des méthodes de groupement, appelées aussi méthodes de coalescence (plus connues selon la terminologie anglaise *clustering* [11]). Le principe est de diviser un ensemble d'échantillons d'apprentissage en sous-ensembles homogènes appelés classes. Parmi les méthodes de *clustering* les plus utilisées, on trouve celles des « nuées dynamiques » [10]. Cette méthode itérative, appelée aussi classification stricte, regroupe les échantillons d'apprentissage de façon à minimiser une erreur au sens des moindres carrées selon une norme donnée [13]. L'inconvénient

majeur de la méthode réside dans le fait que les résultats obtenus peuvent être biaisés par la présence en trop grand nombre de pixels mixtes, c-a-d appartenant à plusieurs classes [12], [13]. C'est pour en tenir compte que des méthodes de classification floue ont été développées dont l'avantage majeur est la prise en considération des échantillons mixtes [12], [13], [17]-[20]. L'idée est de représenter la similarité d'un échantillon avec chaque classe par une fonction dite d'appartenance à valeurs réelles entre 0 et 1. Chaque échantillon est donc membre de chaque classe avec un degré d'appartenance ou de similarité variant de 0 à 1, où la valeur 0 indique l'absence totale de similarité et la valeur 1 la similarité parfaite. La classification stricte est donc un cas particulier de la classification floue si la similarité d'un échantillon avec chaque classe est représentée par une fonction d'appartenance à une seule valeur entière 0 ou 1.

Aussi bien pour la version stricte que la version floue de la méthode *clustering*, les deux versions conduisent à des optimums locaux de la solution du problème de partition de l'ensemble global des échantillons d'apprentissage. En fait, les deux versions se basent sur la minimisation d'une fonction objective au sens des moindres carrées. Cette minimisation n'est que locale puisqu'elle dépend d'une part d'un nombre de classes fixe et d'autre part de l'initialisation des centres de ces classes. D'autant plus que l'algorithme itératif utilisé dans les deux versions est similaire aux algorithmes de type gradient descendant qui fournissent une optimisation locale [16]. Cette optimisation locale pose le problème de validation des groupes obtenus après la classification [15], [16], [21]-[25].

Néanmoins, il est possible de se rapprocher de l'optimum global en utilisant la méthode d'optimisation recuit simulé [26], [27]. Dans ce contexte, un algorithme de groupement flou, optimal est proposé dans notre article [15]. En partant d'une initialisation aléatoire du nombre de classes et des ses centres, l'idée de base est de construire une fonction objective floue dont le minimum global caractérise la bonne partition de l'ensemble des échantillons d'apprentissage au sens de minimisation de l'erreur aux moindres carrées. Afin d'atteindre ce minimum global, la méthode d'optimisation recuit simulé est utilisée. Le nombre de groupes qui mène à une classification robuste est trouvé en comparant entre deux résultats de classifications différentes. Ces deux résultats sont obtenus à partir de deux nombres de groupes différents. La comparaison consiste à trouver la correspondance entre les groupes des deux classifications sans se référer aux réalités de terrain. Une telle comparaison est considérée comme un problème d'optimisation qui sera résolu en utilisant une nouvelle technique d'optimisation basée sur l'analyse de correspondance. La technique est inspirée de la méthode recuit simulé. Les résultats obtenus montrent que l'erreur des moindres carrées peut être réduite par rapport au meilleur résultat obtenu par les deux versions de *clustering*. Cependant, dès que l'ensemble global des échantillons d'apprentissage a une taille plus ou moins importante, le temps de calcul devient prohibitif.

Autre inconvénient de l'algorithme proposé dans [15] est l'aspect empirique dans la mise en œuvre. Le critère de validation des groupes, utilisé dans cet algorithme, est basé sur le calcul du nombre de pixels appartenant à un même groupe dans les deux résultats de classification mises en correspondance. Selon un seuil fixé pour ce nombre, le groupe en considération est retenu ou éliminé. Cependant, le choix de la valeur du seuil est empirique. Si la valeur n'est pas petite, le risque de ne pas détecter les classes, qui sont très étroites telles que les routes et les fleuves, est très élevé. Dans le cas contraire, si la valeur du seuil est assez petite, le résultat de l'analyse de correspondance peut être loin de la réalité du terrain. Cet aspect empirique dans le choix du seuil rend les performances de l'algorithme fortement dépendantes de la valeur du seuil. Pratiquement, le choix adéquat du seuil devient alors une tâche fastidieuse.

Un critère de validation des groupes, plus général et plus efficace, est présenté dans notre article [16]. Le critère est formulé dans un cadre mathématique rigoureux en faisant appel aux méthodes récentes de combinaison de classifieurs [28]-[30]. Nous avons utilisé l'approche Bayésienne de combinaison qui exploite toutes les informations relatives aux classifieurs dans l'élaboration de la règle de combinaison. En considérant les deux résultats de classification comme deux classifieurs qui répondent différemment aux mêmes données d'entrée, en fournissant deux nombres de groupes différents, la combinaison de ces deux classifieurs est effectuée dans le but d'obtenir un résultat de classification plus précis. Un tel résultat est une sorte de validation des groupes puisque certains groupes seront éliminés et d'autres seront retenus. La méthode a été testée et évaluée en utilisant des images POLSAR et les résultats obtenus sont en général satisfaisants. La méthode demande, cependant, beaucoup de paramètres statistiques à estimer et par conséquent un ensemble d'échantillons d'apprentissage représentatif et de taille importante est nécessaire. Le concept théorique de la méthode peut être révisé dans l'optique de réduire au minimum le nombre de paramètres statistiques à estimer et rendre la mise en œuvre et l'application de la méthode plus pratique.

Récemment, une approche similaire à notre méthode proposée dans [16] a été suggérée dans [31] pour une fusion non-supervisée des images hyperspectrales. L'approche est une sorte de combinaison de plusieurs classifications non-supervisées des sous-ensembles des bandes spectrales différents préalablement sélectionnés. Les résultats des classifications sont alors considérés comme des classifieurs différents qui fournissent des nombres de groupes différents. Ces résultats sont ensuite combinés en utilisant la méthode de combinaison de vote majoritaire et ses variantes [28]-[30]. Les résultats de l'approche, en utilisant des données hyperspectrales réelles du système AVIRIS, semblent satisfaisants. Cependant, le concept théorique de l'approche nécessite plus de développement afin de rendre la mise en œuvre de l'approche plus claire et pratique.

2. 2. Avantages et Inconvénients de la Classification Non-Supervisée

Dans les différentes méthodes de classification non-supervisée, l'interprétation des classes obtenues n'est effectuée qu'*a posteriori*, de façon supervisée, par comparaison avec des bases de données ou avec la réalité du terrain. Dans ce sens, on peut dire que les méthodes non-supervisées présentent quelques avantages par rapport aux méthodes supervisées :

- ✓ Une estimation précise des caractéristiques statistiques des classes à partir de zones d'apprentissage est fastidieuse et longue. Le premier intérêt des méthodes non-supervisées est donc leur caractère automatique, particulièrement appréciable dans le cas d'un nombre important de classes.
- ✓ La classification est effectuée selon des critères purement mathématiques indépendants de l'application considérée.
- ✓ Si les méthodes supervisées sont capables de prédire l'existence de classes *a priori* inconnues, par l'introduction d'une classe de rejet, elles sont généralement incapables de séparer les classes inconnues *a priori* et d'estimer leurs caractéristiques. Les méthodes non-supervisées permettront la séparation et l'estimation des caractéristiques des classes inconnues.
- ✓ L'avantage d'opérer de manière non-supervisée est que les images sont classifiées selon l'information réellement présente dans ces images. Les classes observées dans les résultats ne sont pas biaisées par des informations extérieures introduites *a priori*. Ainsi, si deux classes ne sont pas séparables au vu de l'information présente sur une image, la classification non-supervisée ne forcera pas leur séparation.

Cependant, les méthodes non-supervisées présentent aussi quelques inconvénients :

- ✓ Elles sont connues pour être plus coûteuses en temps de calcul.
- ✓ Leur nombre de classes est généralement supérieur à celui des méthodes supervisées. Ceci peut conduire, dans certains cas d'applications, à des présences de classes fictives n'ayant aucun lien avec la réalité terrain.
- ✓ Autre problème crucial, non résolu, et général aux méthodes non-supervisées est l'estimation du nombre optimal de classes qui donne le meilleur résultat de classification. Des approches pour résoudre ce problème ont déjà été proposées [15], [16], [21]-[25]. Néanmoins, ces approches ne fournissent pas des solutions robustes et finales pour ce problème. A notre avis, trouver le nombre de groupes qui fournit le meilleur résultat de la classification est toujours une issue ouverte.

3. Classification Supervisée

Dans la classification supervisée, on utilise des informations *a priori* sur la scène qui peuvent être données sous forme de vérité de terrain. Une mise en correspondance est alors établie entre les données images à classifier et les informations *a priori*. Nous schématisons les étapes d'une classification supervisée par la figure 1. Dans le premier sous-paragraphe, nous allons donner la méthodologie suivie dans la construction du classifieur supervisé. La règle de décision Bayésienne optimale est déduite. Trois méthodes de classification Bayésienne ponctuelle seront exposées en détail à savoir : Classification de maximum de vraisemblances, classification de minimum de distance de Mahalanobis et classification de minimum de distance Euclidienne. Le deuxième sous-paragraphe est consacré à l'apprentissage du classifieur et son influence sur ces performances globales afin de faire ressortir la problématique de la classification supervisée.

3. 1. Construction du Classifieur Supervisé

Une classification consiste à attribuer à chaque pixel x de l'espace de mesure X (vecteur d'images à classifier) une classe $w(x) \in W$ où $W = \{w_1, w_2, \dots, w_M\}$ indique l'ensemble de toutes les classes possibles, où M est le nombre de classes. Souvent, on donne au classifieur la possibilité de ne pas prendre une décision s'il considère que les informations mises à sa disposition ne lui permettent pas de donner un résultat fiable. Ce refus de prise de décision est appelé rejet de la forme x . Une classe de rejet, notée w_{M+1} , est alors rajoutée à l'ensemble de toutes les classes possibles. Soit $W_+ = W \cup \{w_{M+1}\}$, le nouveau ensemble des classes possibles.

Une classification peut donc être considérée comme une fonction de l'espace des formes (pixels) x vers l'espace des classes Cl telle que : $Cl : X \rightarrow W_+$. Elle divise l'espace X en $M + 1$ régions disjointes, qui ne sont pas obligatoirement connexes, appelées régions de décision, R_{w_i} , telle que :

$$w(x) \in w_i \equiv x \in R_{w_i} \quad (1)$$

avec $i = 1, 2, \dots, M$, et $w(x)$ le résultat de classification de la forme x . Pour construire un critère d'évaluation de classification, un coût est attribué à chaque décision prise par le classifieur. Soit w_{vraie} la vraie classe de la forme x , le coût de classification de la forme x comme w_i alors que la vraie classe est w_j est tel que :

$$C(w_i | w_j) = C(w(x) = w_i | w_{vraie} = w_j) \quad (2)$$

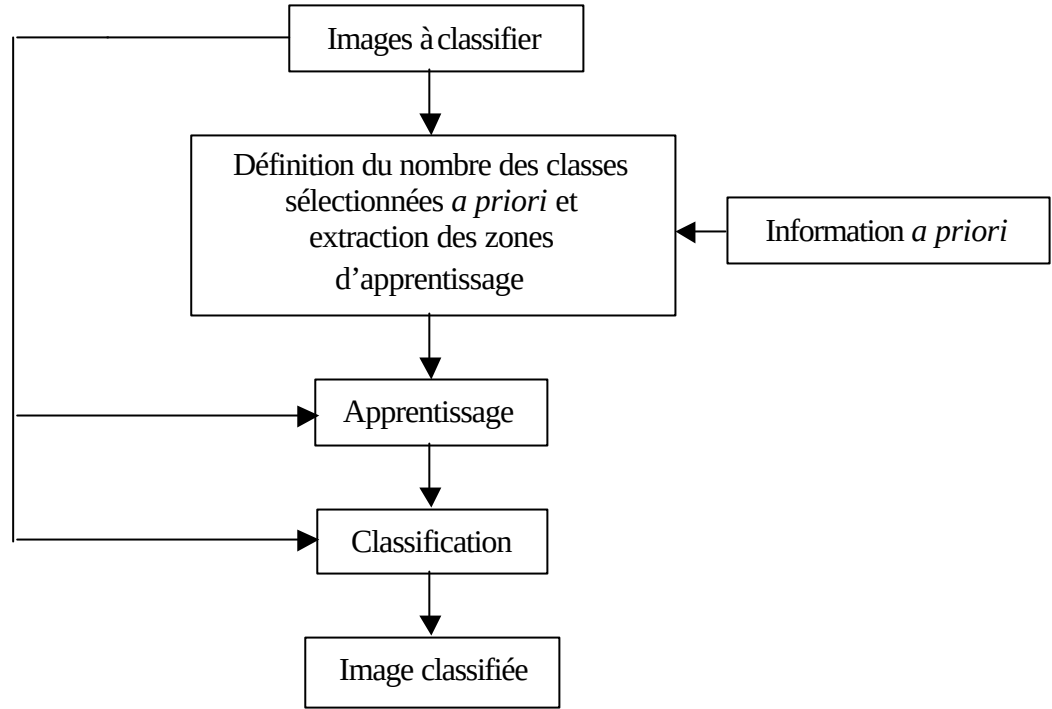


Figure 1. Les étapes d'une classification supervisée.

Au sens de la théorie Bayésienne de décision, les régions de décision, R_{w_i} , sont estimées grâce à la minimisation d'une fonction de risque de nature probabiliste. Le but est de minimiser le coût « moyenne » total de la classification. Ainsi, la décision optimale Bayésienne, appelée aussi la règle de décision de Bayes, consiste à minimiser le coût « moyenne » total pour la forme x de sorte que :

$$w_j(x) = w_{vraie} \text{ si } P(w_j|x) = \min_{w_i \in W_+} (1 - P(w_i|x)) \quad (3)$$

où $P(w_j|x)$ est la probabilité *a posteriori* de la classe w_j . Le coût correspondant à cette décision est appelé le risque de Bayes noté par :

$$R_{Bayes} = \int_X \min_{w_i \in W_+} (P(w_i \neq w_{vraie}|x)) P(x) dx \quad (4)$$

où $P(x)$ est la densité de probabilité de x . Le coût obtenu par la règle de Bayes est le minimum qu'on puisse obtenir et aucune autre méthode ne pourra dépasser ces performances théoriques. Cependant, l'évaluation pratique du risque de Bayes donné en (4) est très difficile. Néanmoins, ce risque peut être reformulé autrement en utilisant la fonction de coût 0-1. Soit $I(i, j)$ la fonction de coût 0-1. Le risque de Bayes conditionnel prend la forme suivante en utilisant la fonction $I(i, j)$:

$$\mathbf{x}_i(x) = \sum_{j=1}^M I(i, j) P(w_j | x) \quad (5)$$

En appliquant la formule de Bayes dans (5), nous obtenons l'expression suivante :

$$\mathbf{x}_i(x) = \sum_{j=1}^M \frac{I(i, j) P(x | w_j) P(w_j)}{P(x)} \quad (6)$$

où $P(w_j)$ est la probabilité *a priori* de la classe w_j . Comme les régions de décision, R_{w_i} , sont exclusives, minimiser le risque conditionnel donné dans (6) revient alors à choisir les régions de décision, R_{w_i} , de sorte que :

$$x \in R_{w_i} \quad \text{si} \quad \mathbf{x}_i(x) < \mathbf{x}_k(x) \quad \text{pour tout} \quad i \neq k \quad (7)$$

$P(x)$ est commune à toutes les classes, on peut donc la supprimer de l'expression du risque conditionnel donnée dans (6). Ainsi, on obtient une nouvelle expression pour le risque conditionnel et l'expression (7) devient :

$$x \in R_{w_i} \quad \text{si} \quad \sum_{j=1}^M I(i, j) P(x | w_j) P(w_j) < \sum_{j=1}^M I(k, j) P(x | w_j) P(w_j) \quad \text{pour tout} \quad i \neq k \quad (8)$$

La classification basée sur la règle de décision donnée dans l'expression (8) est appelée classification Bayésienne optimale. Plusieurs types de classification Bayésienne ont été reportés dans la littérature.

Si chaque pixel est classé à partir de sa valeur, ce type de classification est dit ponctuel. Si chaque pixel est classé en tenant compte des valeurs des pixels contenus dans son voisinage spatial (une notion que sera définie dans le chapitre suivant), il s'agit d'une classification contextuelle. Si chaque pixel est classé à partir de l'ensemble entier des pixels de l'image, cette classification est dite classification de maximum des marginales *a posteriori* "*Maximum of the Posteriori Marginals (MPM)*", dont la solution a été apportée par Marroquin *et al.* [32]. Si toute l'image est classée à partir de l'ensemble entier des pixels de l'image, on a alors la classification qui est basée sur le critère d'optimalité MAP "*Maximum A Posteriori*", qui repose sur la connaissance de la loi conditionnelle de l'image des données observées sachant l'image des classes (ou labels), et de la loi *a priori* de l'image des labels. Ces deux lois peuvent être estimées en utilisant la modélisation Markovienne (qui sera abordée dans le chapitre suivant). Les deux dernières méthodes de classification Bayésienne, MPM et MAP, sont dites globales au sens où elles prennent en compte la totalité de l'information observée, l'ensemble entier des pixels de l'image, par opposition aux deux premières méthodes qui ne sont que locales. Cependant, les classifications ponctuelles peuvent s'interpréter comme une localisation de la classification MPM. Elles représentent le cas extrême où la localisation est réduite au pixel lui-même. La

distinction entre MPM et MAP se traduit par des expressions différentes de la fonction de coût $I(i, j)$. Dans le cas du MAP, cette fonction du coût est de type 0-1. Dans le cas du MPM, le coût est estimé par la somme des pixels mal classifiés. Nous allons donner dans ce qui suit les trois méthodes de classification ponctuelles les plus utilisées en traitement d'images de télédétection.

3. 1. 1. Classification par le maximum de vraisemblances

Soit la fonction de coût $I(i, j)$ de forme suivante :

$$I(i, j) = 1 - f_{ij} \quad (9)$$

avec $f_{ii} = 1$ et f_{ij} à définir. Les risques de Bayes donnés dans (5) et (6) peuvent être reformulés comme suit :

$$\mathbf{x}_i(x) = \sum_{j=1}^M P(w_j|x) - \sum_{j=1}^M f_{ij}.P(w_j|x) = P(x) - g_i(x) \quad (10)$$

$$\text{avec : } g_i(x) = \sum_{j=1}^M f_{ij}.P(w_j|x) \quad (11)$$

$P(x)$ est commune à toutes les classes, et n'aide pas à la discrimination, donc on peut la supprimer de l'expression de $\mathbf{x}_i(x)$ et on obtient :

$$\mathbf{x}_i(x) = -g_i(x) \quad (12)$$

Il s'agit ici de construire le classifieur en utilisant le formalisme des fonctions discriminantes des classes, notées $g_i(x)$ avec $i = 1, 2, \dots, M$, telle que la classe attribuée à une forme x est celle dont la fonction discriminante est la plus grande calculée pour x :

$$w(x) = w_i \text{ si } g_i(x) = \max_{1 \leq j \leq M} g_j(x) \quad (13)$$

Ainsi, si la fonction f_{ij} est égale à la fonction delta de Kroneker telle que :

$$f_{ij} = d_{ij} = \begin{cases} 1 & \text{pour } i = j \\ 0 & \text{pour } i \neq j \end{cases},$$

La classification optimale au sens de Bayes peut être vue comme une classification avec les fonctions discriminantes de type :

$$g_i(x) = P(x|w_i)P(w_i) \quad (14)$$

Ce qui montre d'après (12) qu'il faut choisir les régions de décision, R_{w_i} , qui maximisent la quantité $P(x|w_i)P(w_i)$ sur l'ensemble des indices i :

$$x \in R_{w_i} \text{ si } P(x|w_i)P(w_i) > P(x|w_j)P(w_j) \text{ pour tout } j \neq i \quad (15)$$

Autrement dit on a :

$$x \in R_{w_i} \text{ si } P(x|w_i)/P(x|w_j) > P(w_j)/P(w_i) \text{ pour tout } j \neq i \quad (16)$$

La quantité à gauche de l'inégalité (16) s'appelle le rapport de vraisemblances entre la classe i et toute autre classe j . Nous classons donc la forme x dans la classe i si tous les rapports de vraisemblances entre la classe i et toute autre classe j sont maximisés par rapport à la quantité $P(w_j)/P(w_i)$ (ou 1 si les classes sont équiprobables). La règle de décision (15) est appelée alors règle de décision de maximum de vraisemblances.

Puisque, pour le résultat de la classification, c'est la comparaison des valeurs des fonctions discriminantes qui importe et non pas les valeurs elles-mêmes, cela nous donne la possibilité de remplacer $g_i(x)$ de l'expression (14) par n'importe quelle fonction monotone strictement croissante de g_i sans que le résultat de la classification change. La fonction logarithmique est une fonction croissante et monotone. Dans notre implémentation de la règle de décision de maximum de vraisemblances, nous avons choisi comme fonction discriminante l'expression logarithmique du produit $P(x|w_i)P(w_i)$ afin de n'avoir que des additions dans le calcul des $g_i(x)$. Nous obtenons alors :

$$g_i(x) = \log\{P(x|w_i)\} + \log\{P(w_i)\} \quad (17)$$

et la règle de décision de maximum de vraisemblances devient :

$$x \in w_i \text{ si } g_i(x) > g_j(x) \text{ pour tout } j \neq i \quad (18)$$

Pour chaque classe, nous devons alors estimer à partir des échantillons d'apprentissage les paramètres statistiques de la distribution $P(x|w_i)$, en supposant que les probabilités *a priori*, $P(w_i)$, sont connues. Notons que, lorsque les densités $P(x|w_i)$ sont obtenues par estimation, les probabilités *a priori* $P(w_i)$ peuvent être estimés par des procédures spécifiques [9] si elles sont aussi des inconnues ou si l'hypothèse d'équiprobabilité des classes n'est pas vérifiée.

Cependant, le problème qui peut se poser avec la règle de décision (18) est le cas où on a $g_i(x) = g_j(x)$ pour $i \neq j$. Il s'agit ici d'une surface discriminante où la décision est difficile à prendre. La notion de seuillage est alors introduite dans la règle décision, en

réservant le droit de ne pas affecter la forme x mais de la classer dans la classe de rejet. Ainsi, en incorporant le seuil dans la classification, la règle de décision (18) devient :

$$\begin{array}{ll} \text{si} & g_i(x) > g_j(x) \text{ et } g_i(x) > T_i \quad \text{pour tout } j \neq i \quad x \in w_i \\ \text{si non} & x \in w_{I+M} \end{array} \quad (19)$$

La règle (19) est appelée la règle de maximum de vraisemblances quantifiée où T_i est le seuil qui correspond à la classe w_i , appelé aussi le coût de rejet. Si ce coût est important, il n'y aura jamais de rejet. Cependant, on peut calculer la valeur maximum de T_i pour laquelle il est possible d'avoir des rejets. En effet :

$$\sum_{i=1}^M P(w_i|x) = 1 \Rightarrow g_i(x) = \max_{1 \leq j \leq M} g_j(x) > 1/M \quad (20)$$

Pour que le seuil ait un effet dans la règle de décision (19), il faut que :

$$T_i \leq (1 - M)/M \quad (21)$$

Pour toute valeur de T_i supérieure à cette limite, le rejet n'a aucun effet sur la règle de décision (19). La règle (18) peut être considérée comme un cas particulier de la règle (19). Cependant, l'inconvénient des deux règles est que dans le cas où on aurait plusieurs classes qui ont des fonctions discriminantes égales au maximum et dépassant la valeur du seuil, c.-à-d., $T_i \leq (1 - M)/M$, ces deux règles ne pourront choisir qu'une seule classe parmi ces classes et qui peut ne pas être la vraie classe de la forme x . Dans ces conditions, afin de minimiser l'erreur de classification, on peut utiliser une autre règle dite règle de maximum de vraisemblances notoire telle que :

$$\begin{array}{ll} \text{si} & g_i(x) > g_j(x) \quad \text{pour tout } j \neq i \\ \text{et} & g_i(x) > g_k(x) \quad \text{pour tout } k \neq l \text{ avec } k \neq i \\ \text{et} & g_i(x) - g_k(x) > T_i \\ \text{alors} & x \in w_i \\ \text{si non} & x \in w_{I+M} \end{array} \quad (22)$$

La règle (22) prend en compte seulement la deuxième classe, w_i , qui vérifie le maximum des fonctions discriminantes après la première classe w_i . Afin de prendre en considération l'opposition de toutes les classes et non pas seulement la deuxième classe, une extension de la règle (22), dite règle de maximum de vraisemblances avec opposition restreinte, peut être utilisée :

$$\begin{array}{ll} \text{si} & g_i(x) > g_j(x) \quad \text{pour tout } j \neq i \\ \text{et} & g_i(x) - \sum_{k \neq i} g_k(x) > T_i \\ \text{alors} & x \in w_i \\ \text{si non} & x \in w_{I+M} \end{array} \quad (23)$$

On peut considérer comme fiable une classification quand toutes les classes à l'exception de celle qui vérifie le maximum de fonctions discriminantes sont en deçà du seuil $T_i \leq (1 - M)/M$. Pour cela, il suffit d'examiner la deuxième classe qui vérifie le maximum et on aura la règle suivante, dite règle de maximum de vraisemblances avec opposants faibles :

$$\begin{aligned}
 &\text{si} \quad g_i(x) > g_j(x) \quad \text{pour tout } j \neq i \\
 &\text{et} \quad \max_{k \neq i} g_k(x) < T_i \\
 &\text{alors} \quad x \in w_i \\
 &\text{si non} \quad x \in w_{I+M}
 \end{aligned} \tag{24}$$

On peut aussi combiner deux ou plusieurs de ces règles de décision, afin d'obtenir une meilleure estimation de la vraie classe du pixel x et une classification plus fiable.

3. 1. 2. Classification par le minimum de distance de Mahalanobis

Les fonctions discriminantes, $g_i(x)$, utilisées dans la construction du classifieur peuvent aussi être des distances mesurant la proximité, statistique ou géométrique, de la forme x par rapport à chaque classe candidate. La distance utilisée peut être n'importe quelle norme au sens mathématique du terme.

Soit à considérer la fonction discriminante comme étant égale à la norme quadratique suivante :

$$g_i(x) = (x - m_i)^T \cdot S_i^{-1} (x - m_i) \tag{25}$$

où m_i et S_i sont le vecteur moyenne et la matrice de covariance de la classe w_i . L'expression (25) est une sorte de distance quadratique entre la forme x et le centre de la classe w_i . Cette distance est appelée distance de Mahalanobis [2], [9]. La forme x est alors affectée à la classe w_i si on a :

$$x \in w_i \quad g_i(x) < g_j(x) \quad \text{pour tout } j \neq i \tag{26}$$

L'avantage de cette règle de décision sur la règle de maximum de vraisemblances donnée dans (17) et (18) est que le nombre d'opérations mathématiques à calculer est plus réduit. Le seuil peut être introduit dans la règle de décision de la même manière que dans le cas de la classification par le maximum de vraisemblances. En tenant compte de la valeur du seuil donnée dans (21), et si on inverse le signe de l'expression (25), on a alors les règles de décision (22), (23), et (24) qui peuvent être appliquées en utilisant la norme quadratique.

3. 1. 3. Classification par le minimum de distance Euclidienne :

Une autre norme mathématique plus simple qui peut être utilisée dans la construction du classifieur est la norme Euclidienne. Ainsi, la fonction discriminante, $g_i(x)$, prend la forme suivante :

$$g_i(x) = (x - m_i)^T (x - m_i) \quad (27)$$

Cette fonction discriminante est appelée la distance Euclidienne et la forme x est affectée à la classe dont la distance est la plus petite. La règle de décision (26) reste valable. Cette règle est plus facile à implémenter que la règle de maximum de vraisemblances et la règle de minimum de distance de Mahalanobis. Sachant la valeur du seuil donnée dans (21), et en inversant le signe de la fonction discriminante (27), les règles de décision avec rejet données dans (22), (23), et (24) sont valables en utilisant la norme Euclidienne.

3. 2. Apprentissage dans la Classification Supervisée

La classification par le maximum de vraisemblances s'obtient en choisissant la classe qui donne le maximum de $P(w_j|x)$. Il faut donc connaître les probabilités *a posteriori* de toutes les classes en utilisant la formule de Bayes. Malheureusement, les quantités $P(x|w_j)$ et $P(w_j)$, probabilité conditionnelle et probabilité *a priori*, respectivement, de la classe w_j , qui entrent dans le calcul de $P(w_j|x)$ ne sont presque jamais connues dans la pratique. Tout ce que l'on peut faire est d'utiliser un ensemble d'échantillons tirés de chaque classe pour les estimer.

Soit $X_{w_j}^A$ l'ensemble des échantillons tirés de la classe w_j . En imposant une forme paramétrique pour la fonction de densité de probabilité $P(x|w_j)$, cette dernière est complètement déterminée en estimant les paramètres de la densité à partir de l'ensemble $X_{w_j}^A$ et en utilisant une méthode statistique d'estimation [12].

La probabilité *a priori* $P(w_j)$ est relativement simple à estimer et on peut utiliser une estimation directe sur la base des échantillons d'apprentissage, notée par X^A , et constituée de tous les échantillons tirés des classes, telle que :

$$P(w_j) = \left| \{x \in X^A \mid w_{raie} = w_j\} \right| / |X^A| = |X_{w_j}^A| / |X^A| \quad (28)$$

où $|X_{w_j}^A|$ désigne le nombre d'échantillons de la classe w_j dans l'ensemble global des échantillons d'apprentissage $|X_A|$. D'autres procédures plus sophistiquées, qui ne seront pas données ici, peuvent être utilisées pour estimer $P(w_j)$ [9].

Pour le classifieur utilisant les distances de Mahalanobis, le processus d'apprentissage consiste à estimer les moyennes statistiques et les matrices de covariance des classes. Quant au classifieur utilisant la distance Euclidienne, l'apprentissage se réduit à estimer seulement les moyennes statistiques des classes.

Ce processus d'estimation de $P(x|w_j)$ et $P(w_j)$, pour la classification par le maximum de vraisemblances, et les statistiques d'ordre un et deux, pour la classification par le minimum de distance de Mahalanobis et la classification par le minimum de distance Euclidienne, est appelé processus d'apprentissage du classifieur. Ce processus d'apprentissage est une étape caractéristique des méthodes supervisées et est déterminante du résultat de la classification. Nous allons détailler dans le sous-paragraphe suivant, la démarche à suivre dans l'apprentissage supervisé du classifieur. Ensuite, nous abordons la problématique de l'apprentissage supervisé et son influence sur les performances de la classification effectuée.

3. 2. 1. Sélection des échantillons d'apprentissage

A partir des données de référence, qui peuvent être des cartes thématiques déjà réalisées sur la scène ou mesures réalisées sur le terrain, on définit des zones d'apprentissage, appelées aussi parcelles, qui sont des zones de l'image que l'on juge représentatives d'une certaine classe. Comme nous avons déjà mentionné dans le sous paragraphe 2.1.1 du chapitre 2, deux facteurs essentiels doivent être pris en considération dans le choix des zones d'apprentissage, à savoir :

- 1- La répartition spatiale des zones d'apprentissage d'une classe sur toute la scène à classifier. Si deux régions distinctes spatialement représentent la même classe, il faudrait extraire une parcelle (ou plusieurs) de chaque région. Ceci traduit au mieux la variabilité de la classe. Autre raison principale de la répartition spatiale des zones d'apprentissage est la prise en considération de toutes les classes spectrales qui forment la classe d'information (voir chapitre 2, sous paragraphe 2.1.1 pour les définitions des ces deux termes des classes).
- 2- Extraction d'un nombre suffisamment grand des échantillons d'apprentissage afin d'obtenir des estimateurs précis des paramètres statistiques des classes.

La qualité des échantillons d'apprentissage est un autre facteur à prendre en considération. Les zones d'apprentissage se présentent sous formes de polygones. Donc, on s'attachera tout au moins à isoler, à l'intérieur du contour, les pixels les plus homogènes possibles en éliminant les pixels de bordure.

3. 2. 2. Phase d'apprentissage

La phase d'apprentissage consiste, pour une classe donnée, à rassembler et caractériser l'ensemble des pixels situés dans les zones d'apprentissage sélectionnées. Cette caractérisation nécessite une hypothèse sur la loi de distribution de probabilité de la classe. Ainsi, pour les images multispectrales optiques, la loi Gaussienne est utilisée et les échantillons d'apprentissage sont utilisés pour estimer le vecteur moyenne et la matrice de covariance de la classe. Les différents estimateurs de ces deux statistiques sont détaillés dans le chapitre 2, aussi bien que pour les images multispectrales de dimension faible que pour celles de dimension élevée. Pour les images radars, la distribution K est généralement utilisée (voir le paragraphe 5 du chapitre 2) et les échantillons d'apprentissage sont utilisés pour estimer le paramètre α de cette distribution. Les différentes méthodes d'estimation de ce paramètre ainsi que d'autres aspects importants dans la sélection des échantillons d'apprentissage sont détaillés dans notre publication [33], dont la version intégrale est donnée dans l'annexe A.

3. 2. 3. Problématique de l'apprentissage supervisé

Le problème majeur dans la classification supervisée est la disponibilité d'un nombre suffisant des échantillons d'apprentissage fiables et représentatifs de toutes les classes de la scène imagée. L'extraction des échantillons d'apprentissage est une tâche très fastidieuse, en particulier lorsqu'on n'a pas une connaissance *a priori* suffisante et crédible sur les types de classes existantes dans la scène imagée.

Les images brutes à classer sont généralement bruitées, par conséquent les échantillons extraits de ces images sont entachées de bruit et ne peuvent pas caractériser d'une manière fiable les densités de probabilité des classes. La qualité des échantillons d'apprentissage influe directement sur l'exactitude des estimateurs des paramètres statistiques des densités de probabilité des classes.

Autre facteur qui influe sur la précision de l'estimation des paramètres est la taille de l'ensemble des échantillons d'apprentissage. Ce problème devient crucial lorsque la dimension des données est élevée. Pour l'estimation de la moyenne statistique des classes, ce facteur n'influe pas trop sur l'exactitude de l'estimateur [3], [5], [6]. En revanche,

l'exactitude de l'estimateur de la matrice de covariance de la classe, qui caractérise ces variations autour de sa moyenne et donc son homogénéité et sa forme géométrique, est fortement dépendant du nombre d'échantillons. Un nombre insuffisant des échantillons peut conduire à une matrice de covariance singulière et/ ou une matrice de covariance qui ne décrivent pas fidèlement le comportement statistique de la classe.

3. 2. 4. Comparaison entre les trois méthodes de classification supervisée

Selon le nombre et le type de paramètres statistiques utilisés par la règle de décision du classifieur, la sensibilité de ce dernier au processus d'apprentissage varie. A cet effet, la précision de la classification par le maximum de vraisemblances est fortement dépendante de la taille et la représentativité de l'ensemble des échantillons d'apprentissage. Alors que la classification par le minimum de distance de Mahalanobis est un peu moins sensible à la fiabilité et au nombre d'échantillons d'apprentissage. La classification par le minimum de distance Euclidienne n'est pas très affectée par la qualité et le nombre d'échantillons d'apprentissage [2], [3], [5].

Cependant, l'avantage de la classification par le maximum de vraisemblances par rapport aux deux autres méthodes de classification supervisée réside dans le fait que la règle de décision de maximum de vraisemblances utilise toutes les informations relatives aux comportements statistiques des différentes classes via les densités de probabilités de ces dernières et leurs probabilités *a priori*. Ces informations sont donc obtenues en estimant les paramètres statistiques des densités de probabilités des classes et les probabilités *a priori* de ces dernières. Tandis que la classification par le minimum de distance de Mahalanobis est basée sur les statistiques d'ordre un et d'ordre deux des classes dans sa règle de décision. Si les classes possèdent la loi Gaussienne, comme le cas des images multispectrales optiques, cette règle peut donner de bons résultats, puisque la matrice de covariance de la classe caractérise la forme géométrique et la variabilité spectrale de cette dernière. Cependant, pour les images radars, où l'hypothèse Gaussienne n'est plus valable [33], les informations relatives aux comportements statistiques des classes se trouvent dans les statistiques d'ordre supérieur. Ces statistiques entrent dans l'estimation des paramètres statistiques de la densité de probabilité de la classe (voir les méthodes d'estimation du paramètre α de la distribution K dans notre publication [33]). Dans ce cas là, la classification par le minimum de distance de Mahalanobis est moins performante que la classification par le maximum de vraisemblances. Autre inconvénient de la classification par le minimum de distance de Mahalanobis est de ne pas utiliser l'information liée aux probabilités *a priori* des classes. Cette information est cruciale pour les classes ayant des formes géométriques étroites, telles que les routes et les fleuves, où les probabilités *a priori* permettent de les considérer au mieux dans la règle de décision.

En ce qui concerne la classification par le minimum de distance Euclidienne, le seul paramètre statistique qui entre dans la règle de décision est la statistique d'ordre un de la classe. Par conséquent, elle présente tous les inconvénients de la classification par le minimum de distance de Mahalanobis par rapport à la classification par le maximum de vraisemblances. D'autant plus, que la classification par le minimum de distance de Mahalanobis présente un avantage par rapport à la classification par le minimum de distance Euclidienne en faisant appel à la matrice de covariance de la classe dans la règle de décision.

Puisque au niveau de chaque pixel de l'image à classer, la fonction discriminante du classifieur est calculée pour toutes les classes, la classification par le maximum de vraisemblances est la méthode la plus coûteuse en temps de calcul et en espace mémoire. En effet, le calcul des densités de probabilités et les probabilités *a priori* des différentes classes doit être effectué à l'avance et les résultats devraient être stockés en mémoire. Le temps de calcul de la classification par le minimum de distance de Mahalanobis augmente quadratiquement avec le nombre d'images d'entrée. Alors que dans la classification de minimum de distance Euclidienne, l'augmentation du temps de calcul est linéaire en fonction du nombre d'images d'entrée [3], [5].

3. 3. Critères d'Evaluation de Classification

Afin de pouvoir évaluer le résultat d'une classification, nous avons besoin de méthodes d'évaluation de classification [34]. Une meilleure façon d'évaluer la précision d'une classification est de comparer le résultat obtenu à la réalité du terrain en utilisant les vérités de terrain. Cependant, dans la pratique, il est très difficile d'avoir ces vérités de terrain. On a alors recours à l'appréciation visuelle qualitative de l'image classifiée suivie d'une certaine appréciation quantitative basée sur des critères d'évaluation mathématiques.

Une appréciation visuellement critique permet un premier jugement sur la qualité de l'image classifiée. Cette appréciation qualitative nous donne déjà une idée sur la qualité de l'image, si elle est nette ou bruitée, et sur la cohérence des thèmes présents dans la scène imagée. Cependant, cette appréciation est subjective et d'autant plus compliquée que le nombre de thèmes présents dans l'image classifiée est élevé. En outre, cette appréciation diffère d'une personne à une autre. Nous avons donc besoin d'autres méthodes plus rigoureuses pour pouvoir repérer les erreurs et quantifier la précision de telle ou telle méthode de classification.

Si l'on dispose d'une réalité de terrain sur l'ensemble de la scène ou sur un ensemble de parcelles uniquement, la matrice de confusion constituerait un critère valable pour évaluer la qualité d'une classification. C'est une matrice établie entre l'image classifiée dont on

veut évaluer la qualité et les vérités de terrain. Elle permet non seulement d'estimer le taux de bonne classification mais également les erreurs d'exclusion (ou d'omission) relatives au classifieur et les erreurs d'inclusion (ou commission) relative à l'utilisateur du résultat de la classification.

La matrice de confusion est une matrice carrée ($M \times M$) dont les éléments représentent le nombre de pixels affectés à chaque classe relative à leur appartenance aux classes des vérités de terrain. On choisit arbitrairement de disposer en colonnes les données de référence et en lignes le résultat de la classification. Une classification idéale, implique une matrice de confusion égale à une matrice identité.

Soient c_{ij} les éléments de la matrice de confusion, c_{i+} la somme totale des éléments en lignes, c_{j+} la somme totale des éléments en colonnes, $N = M \times M$ le nombre total des éléments de la matrice. Par définition on a :

$$N = \sum_{i,j} c_{ij}, \quad c_{i+} = \sum_j c_{ij}, \quad \text{et} \quad c_{j+} = \sum_i c_{ij} \quad (29)$$

A partir d'une matrice de confusion, nous pouvons calculer des indicateurs de précision locaux (calculés pour chaque classe) ou globaux (calculés sur l'ensemble des classes). Nous citons les indicateurs suivants :

- Indicateur de précision du classifieur (*Classifiers's Accuracy*) qui mesure la précision locale. Cette précision est caractérisée par le pourcentage de pixels d'une classe de référence affectés à la même classe par le classifieur. Elle est donnée par :

$$C_A = \left(\frac{c_{ii}}{c_{i+}} \right) 100 \quad (30)$$

La mesure $100\% - C_A$ caractérise donc l'erreur d'omission, ou le pourcentage de pixels d'une classe de référence affectés à d'autres classes par le classifieur.

- Indicateur de précision de l'utilisateur (*User's Accuracy*) qui est une autre mesure locale de la précision. Elle indique le pourcentage de pixels d'une classe issue de la classification correspondant à la même classe dans les données de référence. Elle est donnée par :

$$U_A = \left(\frac{c_{ii}}{c_{j+}} \right) 100 \quad (31)$$

La mesure $100\% - U_A$ est l'erreur de commission qui indique le pourcentage de pixels d'une classe issue de la classification mais qui appartiennent à d'autres classes dans les données de référence.

- Indicateur de précision globale de la classification (*Overall Accuracy*) qui représente le rapport entre le nombre total de pixels correctement classés (les éléments digonaux de la matrice de confusion) et le nombre total de pixels de la matrice de confusion. C'est une mesure globale caractérisant le taux global d'une bonne classification. Elle est donnée par :

$$O_A = \left(\left(\sum_i^M c_{ii} \right) / N \right) 100 \quad (32)$$

L'indicateur de précision globale de la classification est indépendant des précisions locales indiquées par les indicateurs de classifieur et de l'utilisateur. Par conséquent, il n'est pas un critère suffisant pour l'évaluation de la qualité d'une classification. Il risque d'induire en erreur l'utilisateur du résultat de la classification. Un autre critère statistique global est alors nécessaire qui prend en compte la précision locale indiquée par le classifieur et la précision locale indiquée par l'utilisateur. Ce critère est appelé le coefficient KAPPA (ou KHAT) [34].

Le coefficient KAPPA est calculé à partir d'une matrice de confusion établie entre l'image classifiée et la réalité de terrain de toute la scène. Il est donné par :

$$\hat{K} = \frac{N \cdot \sum_{i=1}^M c_{ii} - \sum_{i=1}^M (c_{i+} \cdot c_{j+})}{N^2 - \sum_{i=1}^M (c_{i+} \cdot c_{j+})} \quad (33)$$

Si la matrice de confusion est établie entre l'image classifiée et une réalité de terrain ne représentant que quelques parcelles homogènes de la scène, alors l'expression (33) est celle du critère statistique le plus utilisé qui est le coefficient KHAT. Le coefficient KHAT, contrairement au coefficient KAPPA, n'évalue que la précision de bonne classification des parcelles homogènes définies comme données de référence. D'autres critères statistiques ont été dérivés du KHAT. On cite par exemple, la variance du KHAT et le paramètre statistique Z [34]. Ces deux critères globaux permettent une interprétation similaire à celle donnée par KHAT. Autre critère dérivé de KHAT appelé le KHAT local a été proposé récemment [35] dont l'expression est telle que :

$$\hat{K}_i = \frac{N \cdot c_{ii} - (c_{i+} \cdot c_{j+})}{N \cdot c_{i+} - (c_{i+} \cdot c_{j+})} \quad (34)$$

Les critères d'évaluation de classification que nous venons de décrire sont basés sur la matrice de confusion. Cette dernière est calculée en utilisant les données de référence. Ces données de référence doivent être totalement ou partiellement différentes de celles ayant servi à l'élaboration de l'image classifiée. Si l'on ne dispose pas de vérité de terrain, ce qui est généralement le cas dans la pratique, il faut procéder à une étape de définition d'une réalité de terrain qu'on utilisera ensuite pour établir la matrice de confusion. Les étapes à suivre dans la définition de la réalité de terrain sont les suivantes [36] :

- 1- Partition de l'ensemble global des échantillons d'apprentissage, noté S , en deux sous-ensembles :
 - S_1 : Sous-ensemble des échantillons d'apprentissage pour le classifieurs.
 - S_2 : Sous-ensemble des échantillons test à partir duquel sera constituée la matrice de confusion.
- 2- Calcul du coefficient KHAT global $\hat{K}$.
- 3- Permutation des rôles de chacun des deux sous-ensembles S_1 et S_2 .
- 4- Calcul du nouveau coefficient KHAT global $\hat{K}'$.
- 5- Décision sur la vérité de terrain à choisir : La vérité de terrain sera le sous-ensemble d'apprentissage qui donne le meilleur KHAT.

3. 4. Résultats Expérimentaux

Les trois méthodes de classification supervisée que nous avons exposé ainsi que l'approche de définition d'une réalité de terrain, ont été testées et évaluées sur une image multispectrale réelle. Nous remercions Pr. S. B. Serpico et Dr F. Melgani, de l'Université de Gênes en Italie, d'avoir eu la bonté de nous fournir l'image multispectrale utilisée ici. L'image est composée de six images spectrales, représentées par la figure 2, correspondant aux six bandes spectrales du système de cartographe thématique aéroporté "*Airborne Thematic Mapper (ATM)*". Les six images spectrales, acquises par les six capteurs TM aéroportés sur un avion en juillet 1989, représentent une scène d'étude de 250x350 pixels d'une région agricole située près de Feltwell en Royaume Uni (tous droits réservés à *Hunting Technical Services Ltd. (UK)*). Les champs agricoles sont représentés par cinq classes de la couverture terrestre correspondant, respectivement, à cinq types de récoltes : carottes (classe 1), pommes de terre (classe 2), sol nu (classe 3), chaume (classe 4) et betteraves à sucre (classe 5).

La vérité du terrain que nous avons retenu pour la scène imagée est représentée par l'image de la figure 3. Cette vérité de terrain a été obtenue à partir de la matrice de confusion en modifiant plusieurs fois le nombre et les caractéristiques des parcelles d'apprentissage sélectionnées jusqu'à ce que l'indicateur d'erreurs sur les parcelles d'apprentissage sélectionnées soit minimum. La démarche suivie était d'abord d'effectuer une évaluation qualitative du résultat de chaque méthode de classification. Ensuite, l'évaluation quantitative est effectuée en utilisant la matrice de confusion et le coefficient KHAT [36]. Ainsi, une valeur élevée du KHAT correspond à une bonne homogénéité des zones correspondantes aux données de référence. Le nombre d'échantillons d'apprentissage par classe, ainsi extraits des différentes parcelles sélectionnées définitivement dans les cinq classes constituant la vérité de terrain retenue, sont donnés dans le tableau 1.

Les images résultats issues de la classification par le maximum de vraisemblances, classification par le minimum de distance de Mahalanobis, et classification par le minimum de distance Euclidienne sont données par les figures 4, 5, et 6, respectivement. Une vue globale des trois résultats et une comparaison visuelle de chaque résultat avec les zones d'apprentissage de la figure 3, nous permettent de dire que le meilleur résultat est celui obtenu par la classification par le maximum de vraisemblances. L'image correspondante est plus homogène, peu de pixels mal classifiés et les classes sont bien délimitées. L'image de la classification par le minimum de distance de Mahalanobis est meilleure par rapport à l'image de la classification par le minimum de distance Euclidienne. Cette dernière présente des confusions entre la classe 1 (carottes, couleur verte) et la classe 2 (pommes de terre, couleur blanche) et entre la classe 4 (chaume, couleur bleue) et la classe 5 (betteraves à sucre, couleur noir). L'image de la figure 5 présente aussi des confusions entre ces quatre classes, mais avec un degré moins.

Pour l'évaluation quantitative des performances des trois méthodes de classification, nous avons utilisé les trois critères de performances suivants :

- Taux de reconnaissance : C'est le taux de classification correcte des pixels des zones d'apprentissage, noté t_{rec} .
- Taux d'erreur : C'est le taux de classification incorrecte des pixels des zones d'apprentissage, noté t_{erreur} .
- Taux de rejet : C'est le taux de non-classification (rejet) des pixels des zones d'apprentissage, noté t_{rejet} .

Les zones d'apprentissage sont celles de l'image de réalité de terrain de la figure 3. Pour le taux de rejet, nous avons constaté qu'il n'y a pas de rejet dans la classification de maximum de vraisemblances et qu'il est faible dans les deux autres méthodes. Les figures 7 et 8 donnent, respectivement, les valeurs de taux de reconnaissance et taux d'erreur pour les trois méthodes de classification. Ces valeurs montrent aussi que la méthode de classification de maximum de vraisemblances est la méthode la plus performante.

Nous justifions ces résultats par le fait que les échantillons d'apprentissage ainsi sélectionnés ont été perfectionnés à travers le processus de définition de réalité de terrain. Par conséquent, comme nous avons déjà mentionné dans le sous-paragraphe 3.2.4 ci-dessus, la classification par le maximum de vraisemblances est la plus performante puisqu'elle exploite toutes les informations relatives aux comportements statistiques des classes via les densités de probabilité et les probabilités *a priori* des classes. Le résultat de la classification par le minimum de distance de Mahalanobis est meilleur par rapport au résultat de la classification par le minimum de distance Euclidienne, parce que cette dernière n'utilise que la statistique d'ordre un dans sa règle de décision, alors que la première fait appel aussi à la statistique d'ordre deux dont l'avantage est de caractériser la variabilité spectrale interne de la classe.

4. Classification d'Images Multispectrales de Dimension Élevée

Les méthodes de classification supervisée discutées ci-dessus ont été développées pour les appliquer sur des images multispectrales de dimension faible ou modérée telles que les images fournies par la série de satellites Landsat et la série de satellites SPOT. Ces méthodes ne peuvent pas être appliquées directement pour classer les images multispectrales de dimension élevée fournies par les satellites de la nouvelle génération telles que celles fournies par les systèmes AVIRIS, MODIS et HIRIS “*High Resolution Imaging Spectrometer*”.

Les problèmes rencontrés dans l'application de méthodes de classification classiques sur les images multispectrales de dimension élevée sont les suivantes :

- Le temps de classification augmente de manière significative.
- Les classes peuvent ne pas avoir assez d'échantillons d'apprentissage disponibles pour assurer des estimateurs précis de leurs paramètres statistiques.
- L'importance agrandit des statistiques du premier et second ordre pour la discrimination entre les classes dans les images multispectrales de dimension élevée et donc l'effet des évaluations imprécises de ces statistiques sur le résultat de la classification devient un problème crucial.

Surmonter ces inconvénients est une tâche majeure dans plusieurs études récentes [5]-[8], [37]-[41]. Les méthodes de classification ainsi développées sont généralement des versions modifiées des trois méthodes de classification supervisée que nous avons exposé dans le paragraphe 3. A notre avis, la classification d'images multispectrales de dimension élevée reste encore une issue ouverte en raison de problème d'extraction d'un nombre suffisant d'échantillons d'apprentissage représentatifs qui devient une tâche pratiquement impossible. Nous donnons dans ce qui suit une synthèse des différentes approches et méthodes de classification d'images multispectrales de dimension élevée, proposées dans la littérature, tout en faisant ressortir les inconvénients et les avantages de chacune d'elles.

4. 1. Approches Basées sur la Réduction de la Dimensionnalité

Ces approches sont basées sur la réduction des bandes spectrales ou la sélection de caractéristiques par des transformations linéaires tout en préservant autant d'informations que possible. Les bandes ou les caractéristiques ainsi retenues sont utilisées ensuite comme entrées pour les méthodes de classification classiques telles que les trois méthodes de classification supervisée étudiées précédemment.

Cependant, les transformations linéaires sont appliquées sur un nombre important de bandes spectrales et par conséquent le temps de calcul est important. De plus, la détermination de la matrice de transformation linéaire est souvent en fonction des vecteurs propres des matrices de covariance des classes ou de l'ensemble entier des bandes spectrales. Les estimateurs des matrices de covariance nécessitent un nombre important d'échantillons d'apprentissage. Ceci est un problème crucial qui influe directement sur la précision des résultats des transformations linéaires. La décomposition spectrale de la matrice de covariance présente l'inconvénient d'accumuler les erreurs de calcul dans le cas où la dimension de la matrice est élevée.

4. 2. Classification de Maximum de Vraisemblances à Plusieurs Etapes

La classification de maximum de vraisemblances à plusieurs étapes réduit en général le temps d'exécution en réduisant le nombre de classes considéré à chaque étape de classification. L'idée de la classification de maximum de vraisemblances à trois étapes [37], par exemple, est d'estimer l'intervalle des valeurs du terme quadratique de la forme Gaussienne de la fonction discriminante donnée dans (17), en se basant sur des théorèmes algébriques, et donc éliminer quelques classes dans le processus de recherche de la vraie classe. Les valeurs précises du terme quadratique sont calculées seulement à l'étape finale. Cette méthode cependant n'obvient pas au problème du nombre d'échantillons d'apprentissage nécessaire pour les petites classes.

4. 3. Classification de Maximum de Vraisemblances Rapide

La classification de maximum de vraisemblances rapide [38] utilise une fraction de bandes spectrales à chaque étape ; le nombre de bandes utilisées augmente au fur et à mesure pendant les étapes de classification. A une étape donnée, si une classe a une valeur de vraisemblance inférieure à un seuil fixé, elle est considérée peu probable pour le pixel en considération et est enlevée dans le reste du processus de classification. Finalement, seulement un nombre restreint de classes doivent être considérées à l'étape finale quand toutes les bandes spectrales disponibles sont utilisées. Comme le cas de la classification à plusieurs étages, cette méthode souffre du problème du nombre d'échantillons d'apprentissage nécessaire pour les petites classes.

4. 4. Classification de Maximum de Vraisemblances Modifiée

Une méthode de classification de vraisemblances modifiée a été proposée dans [39], dans laquelle le modèle Gaussien de la classe a été modifié. L'ensemble entier de bandes

spectrales est subdivisé en plusieurs sous-groupes en se basant sur la nature de la corrélation globale entre les différentes bandes. En ignorant la corrélation entre les sous-groupes, ces derniers sont considérés statistiquement indépendants. Par conséquent, chaque classe est modélisée par une distribution Gaussienne dans chaque sous-groupe de bandes spectrales. La fonction discriminante globale de la classe devient la somme logarithmique des fonctions discriminantes des sous-groupes. La classification de maximum de vraisemblances classique et la classification de minimum de distance Euclidienne ont été utilisées pour la comparaison. Les résultats de la classification de maximum de vraisemblances modifiée, obtenus en utilisant des données spécifiques fournies par le système AVIRIS, sont satisfaisants en termes d'exactitude de classification et réduction de temps de calcul. Cependant, ces résultats particuliers n'impliquent pas que l'hypothèse d'indépendance statistique des sous-groupes de bandes spectrales (qui est la base de cette méthode) est toujours vraie.

4. 5. Approches Rapides de Classification de Minimum de Distance de Mahalanobis

Les approches rapides de classification de minimum de distance de Mahalanobis proposées dans [39] sont basées sur des simplifications algébriques de l'expression de la distance de Mahalanobis donnée dans (25), en utilisant des théorèmes de l'algèbre linéaire telles que : le théorème du rang de la forme quadratique [37] ; la méthode de Winograd [41], [42] et l'approche de l'ensemble moyen [43]. L'efficacité des méthodes proposées pour réduire le temps de calcul est donnée par deux paramètres, à savoir, la vitesse théorique et la vitesse expérimentale (observée). La vitesse théorique d'un algorithme est définie par le rapport du nombre de multiplications nécessaires pour le classifieur au nombre de multiplications nécessaires pour classifier un pixel. Les résultats publiés montre l'efficacité des approches proposées dans la réduction de temps de calcul, cependant, la précision des résultats n'a pas été indiquée.

4. 6. Classification de Minimum de Distance Euclidienne Modifiée

Une méthode de classification de minimum de distance Euclidienne modifiée est proposée dans [6] pour classifier les images multispectrales de dimension élevée. La méthode est basée sur deux nouvelles définitions de la moyenne statistique de la classe. La première définition est celle de la différence de la moyenne de la classe "*Average Class Mean Difference (ACMD)*". L'expression d'une telle moyenne est telle que :

$$ACMD = \frac{2}{M(M-1)} \sum_{i=2}^M \sum_{j=1}^{i-1} |m_i - m_j| \quad (35)$$

En utilisant cette nouvelle expression de la moyenne statistique, les performances de la classification de minimum de distance Euclidienne dans le cas d'images multispectrales de dimension élevée, sont améliorées par rapport à la méthode classique.

La deuxième définition de la moyenne statistique de la classe est celle de la moyenne des distances calculées à partir de la moyenne de la classe "*Average Distance From Mean (ADFM)*", dont l'expression est telle que :

$$ADFM = \frac{1}{N} \sum_{i=1}^M \sum_{j=1}^{N_i} |x_j^i - m_i| \quad (36)$$

où N est l'ensemble total d'échantillons d'apprentissage, N_i le nombre d'échantillons de la classe i et x_j^i est le $j^{ième}$ échantillon de la classe i . Si la valeur de ADFM diminue, les performances de la classification de minimum de distance Euclidienne augmentent. Les résultats théoriques relatifs à l'utilisation de ces nouvelles définitions de la moyenne statistique de la classe sont satisfaisants. Néanmoins, le test de cette méthode de classification modifiée sur des images multispectrales réelles est nécessaire pour pouvoir évaluer réellement son efficacité.

6. Conclusion

Dans ce chapitre, nous avons abordé la classification ponctuelle statistique des images de télédétection civile. Les deux catégories de méthodes de classification souvent utilisées dans la télédétection sont les méthodes de classification non supervisée et les méthodes de classification supervisée. Nous avons exposé les avantages et les inconvénients de chacune des deux catégories. Seules les méthodes supervisées ont été considérées en détail. Ainsi, nous avons donné la méthodologie suivie dans la construction d'un classifieur supervisé. Les règles de décision Bayésienne optimales, avec ou sans seuillage, ont été déduites et les différentes méthodes de classification Bayésienne sont données. Trois méthodes ont été détaillées à savoir : la classification de maximum de vraisemblances, la classification de minimum de distance de Mahalanobis et la classification de minimum de distance Euclidienne. Une étude comparative entre ces trois méthodes a été effectuée.

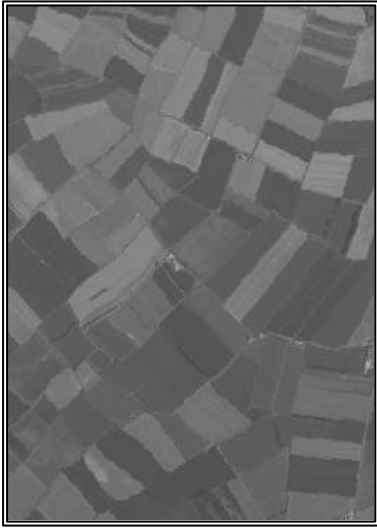
L'apprentissage d'un classifieur supervisé et son influence sur ces performances globales sont abordés en détail afin de faire ressortir la problématique de la classification supervisée. Afin d'évaluer le résultat d'une classification, des critères d'évaluation de la classification sont exposés et une approche de détermination de réalité de terrain est donnée. Cette approche nous a permis d'affranchir le problème d'évaluation de la classification basée sur des critères mathématiques dans les conditions où les vérités de terrain ne sont pas disponibles. Des résultats expérimentaux ont été donnés, en utilisant une image multispectrale optique réelle, où les trois méthodes de classification supervisée ont été testées et évaluées ainsi que l'approche de détermination de réalité de terrain.

La classification d'images multispectrales de dimension élevée reste encore une issue ouverte. Les différentes méthodes proposées dans la littérature sont en général des versions modifiées des trois méthodes de classification supervisée que nous avons détaillé dans ce chapitre. Ainsi, nous avons donné un état de l'art sur les différentes méthodes de classification d'images multispectrales de dimension élevée proposées tout en faisant ressortir les inconvénients et les avantages de chacune.

Références

- [1] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. Mc Graw-Hill, 1978.
- [2] J. A. Richards, *Remote Sensing Digital Image Analysis. An Introduction*, 2nd Edition, Springer-Verlag, Berlin, 1993.
- [3] S. Chitroub, *Implémentation d'un processus d'analyse et de classification d'images satellitaires. Application à la caractérisation des états de surface*. Thèse de Magister, en Traitement D'Images-Télédétection, Institut d'Electronique, USTHB, Alger, 03/94-M/EL, May 1994.
- [4] P. E. Ardanuy, D. Han, and V. V. Salomonson, "The moderate resolution imaging spectrometer (MODIS) science and data system requirements", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 29, pp. 75-88, 1991.
- [5] C. Lee and D. Landgrebe, "Analyzing high-dimensional multispectral data", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 31, No. 4, pp. 792-800, July 1993.
- [6] C. Lee and D. Landgrebe, *Feature Extraction and Classification Algorithms for High Dimensional Data*, Ph.D. Thesis, School of Electrical and Computer Engineering, Purdue University, West Lafayette, TR-EE 93-1, January 1993.
- [7] J. P. Hoffbeck and D. Landgrebe, *Classification of High Dimensional Multispectral Data*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-EE95-14, May 1995.
- [8] Q. Z. Jackson and D. Landgrebe, *Design of an Adaptive Classification Procedure for the Analysis of High-Dimensional Data with Limited Training Samples*. Ph. D. Thesis, School of Electrical Engineering, Purdue University, West Lafayette, TR-ECE 01-5, December 2001.
- [9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.
- [10] G. Celeux, E. Diday, G. Govaert, Y. Lechevallier, et H. Ralambondrainy, *Classification Automatique des Données*, Dunod Informatique, 1989.
- [11] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd Edition, Academic Press, Boston, 1990.
- [12] A. J. Richard and D. W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall International, New-Jersey, 1990.
- [13] S. Le Hégarat-Masclé, *Classification Non Supervisée d'Images SAR Polarimétriques*. Thèse présentée pour obtenir le grade de docteur de l'Ecole Nationale Supérieure des Télécommunications, ENST 96 E 024, septembre 1996.
- [14] S. Chitroub and B. Sansal, "Méthodes de classification non supervisée stricte et floue appliquées aux images de télédétection. Principes et étude comparative", 2^{ème} Conférence Internationale d'Electronique, Signaux, Systèmes et Automatique (SSA2'99), du 10 au 12 mai, 1999, Blida, Algérie, pp. 55-77.
- [15] S. Chitroub, A. Houacine, and B. Sansal, "Robust optimal fuzzy clustering algorithm applicable to multispectral and polarimetric synthetic aperture radar images", *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 21-25, September 1999, Florence, Italy, Vol. 3871, pp. 325-336.
- [16] S. Chitroub, A. Houacine, and B. Sansal, "Bayesian combination method of classifiers for cluster validation problem", *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 21-25, September 2000, Barcelona, Spain, Vol. 3501, pp. 325-336.
- [17] L. A. Zadeh, "Fuzzy sets", *Information and Control*, Vol. 8, pp. 338-358, 1965.

-
- [18] J. C. Bezdek, "A convergence theorem for the fuzzy c-means clustering algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 2, pp. 1-8, 1980.
 - [19] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: fuzzy c-means clustering algorithm", *Computer & Geoscience*, Vol. 10, pp. 191-203, 1984.
 - [20] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering", *IEEE Transactions on Fuzzy Systems*, Vol. 1, No. 2, pp. 98-110, 1993.
 - [21] J. C. Bezdek, "Cluster validity with fuzzy sets", *Journal of Cybernetic*, Vol. 3, No. 3, pp. 58-72, 1974.
 - [22] Y. Fukuyama and M. Sugeno, "A new method for choosing the number of cluster for the fuzzy c-means method", In *Proceedings of 5th Fuzzy Systems Symposium*, Japan, pp. 247-250, 1989.
 - [23] X. L. Xie and G. A. Beni, "Validity measure for fuzzy clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 3, No. 8, pp. 841-846, 1991.
 - [24] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model", *IEEE Transactions on Fuzzy Systems*, Vol. 3, No. 3, pp. 370-379, 1995.
 - [25] R. N. Davé and R. Krishnapuram, "Robust clustering methods: a unified view", *IEEE Transactions on Fuzzy Systems*, Vol. 5, No. 2, pp. 270-292, 1997.
 - [26] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing", *Science*, Vol. 220, pp. 671-680, 1983.
 - [27] P. J. Van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Dordrecht: D. Reidel Publisher Company, Netherlands, 1987.
 - [28] J. Kittler, "Combining classifiers: A theoretical framework", *Pattern Analysis and Application*, Vol. 1, pp. 18-27, 1998.
 - [29] J. Kittler, R. P. W. Duin, M. Hafet, and J. Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 226-239, March 1998.
 - [30] B. Moobed, *Combinaison de Classifieurs, Une Nouvelle Approche*. Thèse pour obtenir le grade de Docteur en sciences de l'université Paris XI Orsay, France, 1996.
 - [31] L. O. Jiménez, M. Velez-Reyes, J. L. Rivera, and H. T. Velazquez, "Unsupervised data fusion for hyperspectral imaging", In *Proceedings of SPIE, Image and Signal Processing for Remote Sensing Conference*, 18-21 September 2001, Toulouse, France, Vol. 4541, pp. 180-189.
 - [32] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision", *Journal of the American Statistical Association*, Vol. 82, pp. 76-89, 1987.
 - [33] S. Chitroub, A. Houacine, and B. Sansal, "Statistical characterisation and modelling of SAR images", *Signal Processing*, Vol. 82, No. 1, pp. 69-92, February 2002.
 - [34] R. G. Congalton, "A review of assessing the accuracy of classification of remotely sensed data", *Remote Sensing of Environment*, Vol. 37, pp. 35-46, 1991.
 - [35] M. A. Shabah and O. Dikshit, "Improvement of classification in urban areas by the use of textural features: The case study of Lucknow city, Uttar Pradesh", *International Journal of Remote Sensing*, Vol. 22, No. 4, pp. 565-593, 2001.
 - [36] R. Khedam, *Etude et Développement de Méthodes de Classification Supervisée Contextuelle d'Images Satellitaires*. Thèse de Magister en Traitement d'Images-Télédétection, Faculté d'Electronique et d'Informatique, USTHB, Alger, novembre 2001.
 - [37] N. B. Venkateswarlu and P. S. V. S. K. Raju, "Three stage maximum likelihood classifier", *Pattern Recognition*, Vol. 24, pp. 1113-1116, 1991.
 - [38] C. Lee and D. Landgrebe, "Fast likelihood classification", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 29, pp. 509-517, 1991.
 - [39] X. Jia and J. A. Richards, "Efficient maximum likelihood classification for imaging spectrometer data sets", *IEEE Transaction on Geoscience & Remote Sensing*, Vol. 32, No. 2, pp. 274-281, 1994.
 - [40] N. B. Venkateswarlu and R. P. Singh, "Various approaches to speed up Mahalanobis distance classifier", *International Journal of Remote Sensing*, Vol. 16, No. 16, pp. 3157-3163, 1995.
 - [41] L. Lydio Kronsjo, *Algorithms: Their Complexity and Efficiency*. John Willey, Singapore, 1987.
 - [42] N. B. Venkateswarlu and P. S. V. S. K. Raju, "Winograd's method: a perspective for some pattern recognition problems", *Pattern Recognition Letters*, Vol. 15, pp. 105-109, 1994.
 - [43] F. W. Zaki, A. I. Fattah, Y. M. Enab, and S. H. E. Konyaly, "An ensemble average classifier for pattern recognition machines", *Pattern Recognition*, Vol. 21, pp. 327-332, 1988.
-



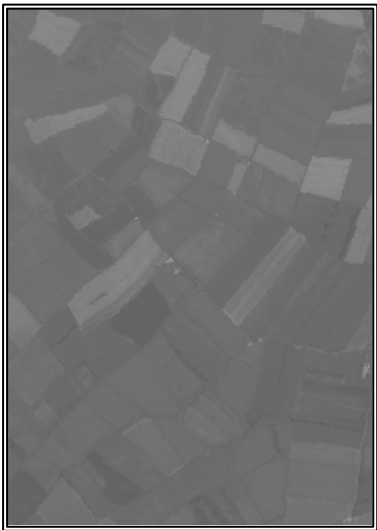
Bande spectrale ATM10.



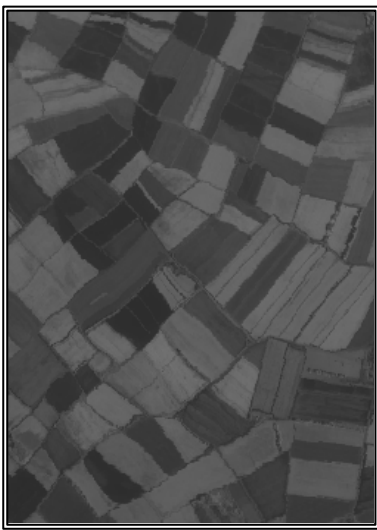
Bande spectrale ATM2.



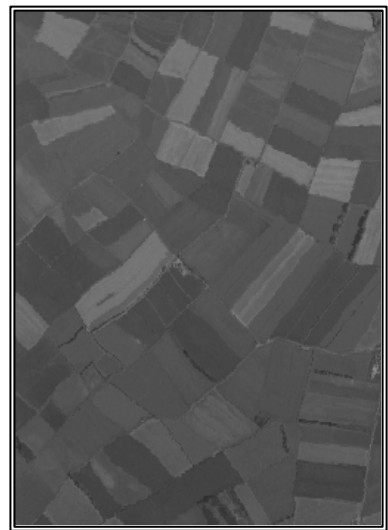
Bande spectrale ATM3.



Bande spectrale ATM5.



Bande spectrale ATM7.



Bande spectrale ATM9.

Figure 2. Les six bandes spectrales du système de cartographe thématique aéroporté “*Airborne Thematic Mapper (ATM)*”. Les six images spectrales, acquises par les six capteurs TM aéroportés sur un avion en juillet 1989, représentent une scène d’étude de 250x350 pixels d’une région agricole située près de Feltwell en Royaume Uni.
(tous droits réservés à *Hunting Technical Services Ltd. (UK)*).

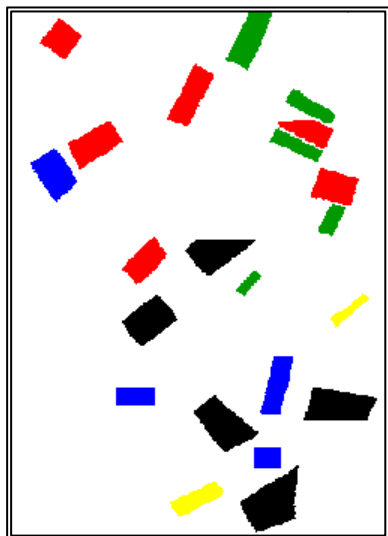


Figure 3. Image de réalité de terrain.
(zones d'apprentissage sélectionnées).

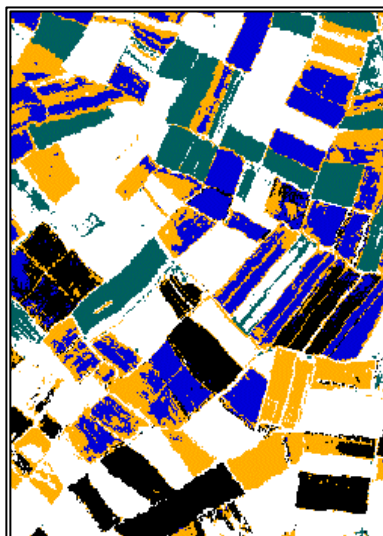


Figure 4. Classification de minimum
de distance Euclidienne.

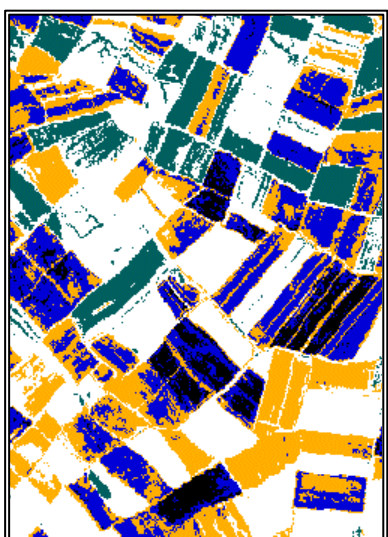


Figure 5. Classification de minimum
de distance de Mahalanobis.

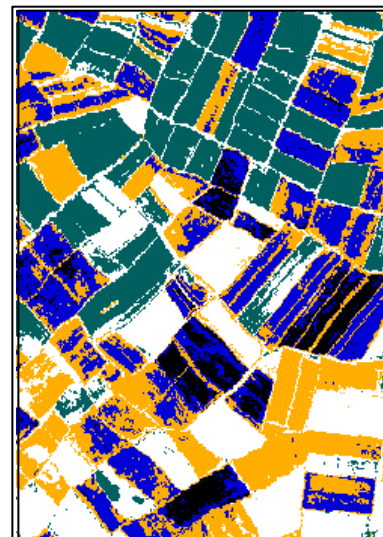


Figure 6. Classification de maximum
de vraisemblances.

Classe	Thème	Couleur dans l'image de réalité de terrain.	Couleur dans les images classifiées.	Nombre de pixels dans les zones d'apprentissage.
Classe 1	Carottes	Rouge	Vert	3015
Classe 2	Pommes de terre	Noir	Blanc	3948
Classe 3	Sol nu	Bleu	Jaune	1685
Classe 4	Chaume	Vert clair	Bleu	1596
Classe 5	Betteraves à sucre	Jaune clair	Noir	559

Tableau 1. Nombre d'échantillons d'apprentissage extraits des différentes parcelles sélectionnées définitivement dans les cinq classes constituant la vérité de terrain retenue.

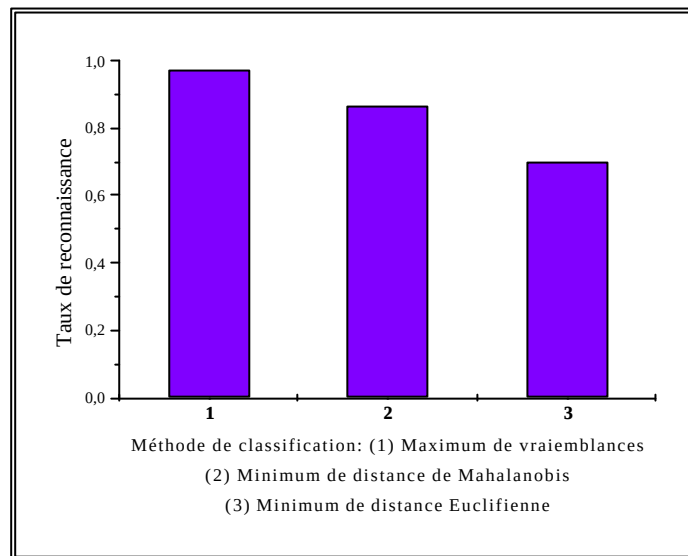


Figure 7. Taux de reconnaissance dans les trois méthodes de classification.

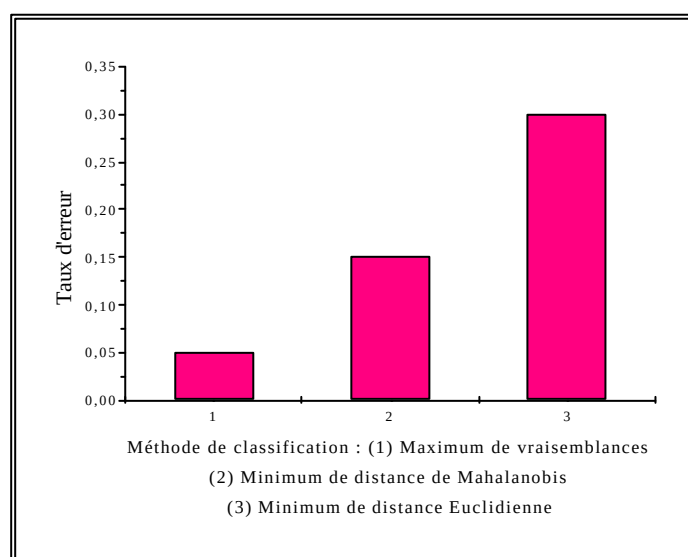


Figure 8. Taux d'erreur dans les trois méthodes de classification.

Chapitre 5

Classification Contextuelle d'Images de Télédétection

1. Introduction

L'inconvénient de la classification ponctuelle réside dans le fait que le pixel de l'image est classé en utilisant seulement sa valeur radiométrique sans tenir compte des valeurs des pixels voisins [1]-[4]. Le résultat est souvent une image bruitée qui présente l'apparence d'un phénomène communément appelé sel et poivre ("*salt and pepper*" en terminologie anglaise) nuit à la lisibilité et à l'interprétation de l'image classifiée.

Les méthodes de classification contextuelle se proposent d'être des méthodes plus robustes dans le sens où elles consistent à optimiser les résultats de classification ponctuelle. Ces méthodes reposent sur la notion de l'information contextuelle (spatiale) apportée à un pixel par un voisinage qui lui est approprié, en utilisant des outils mathématiques rigoureux [6]-[11]. En raison de la capacité dont il dispose pour l'incorporation d'information de contexte, le formalisme de la modélisation du contexte spatial par les champs aléatoires de Markov "*Markov Random Fields (MRF)*" est l'outil le plus répandu. Ainsi, la prise en considération du contexte spatial d'un voisinage donné, s'appuie sur la modélisation Markovienne de l'image et théorie des graphes [12].

Le paragraphe 2 du présent chapitre est consacré à la modélisation Markovienne d'images ainsi qu'aux notions importantes et nécessaires pour l'élaboration d'un modèle Markovien pour la classification contextuelle supervisée. Ainsi, nous traitons le fondement des champs aléatoires de Markov et comment il est possible, grâce au théorème d'*Hammersley-Clifford*, de les spécifier simplement à partir de fonctions de potentiels correspondant à des distributions locales de probabilités [5], [13], [14]. Nous présentons le modèle Markovien dans un cadre général facile à utiliser et implémenter. Nous donnons les différents estimateurs Bayésiens de champ des étiquettes et la manière de les exprimer sous formes énergétiques.

La classification Bayésienne associée à la modélisation Markovienne d'images présente une fonction d'énergie (objective) à minimiser. Le minimum global de cette fonction est la caractéristique de l'image classifiée recherchée [14]. Le formalisme MRF-Bayes ainsi que la fonction d'énergie correspondante sont abordées dans le paragraphe 3. La fonction d'énergie déduite du formalisme MRF-bayes est non-convexe et sa minimisation pose le problème d'optimisation combinatoire. Ainsi, pour la minimiser, nous avons étudié et implémenté les algorithmes d'optimisation combinatoire détaillés dans le paragraphe 4. Les « méta-heuristiques » principalement le recuit simulé [13], [15]-[19], la méthode de recherche tabou [18]-[21], les algorithmes génétiques [18], [19], [22]-[24] sont considérées comme des méthodes efficaces pour la résolution de problèmes d'optimisation combinatoires [18], [19]. Le travail qui nous présentons ici consiste à adapter ces méthodes en vue du traitement des fonctions d'énergie discrètes dont le minimum global caractérise une classification contextuelle optimale de l'image.

2. Modélisation Markovienne d'Images

Dans les images de télédétection spatiale civile, les pixels ne sont pas indépendants les uns des autres. Les méthodes de classification Bayésienne globale dont le principe est la classification de toute l'image (dans le cas de MAP) ou d'un seul pixel de l'image (dans le cas de MPM) à partir de tous les pixels de l'image, devraient *a priori* donner de meilleurs résultats. Cependant, la mise en œuvre de MAP est confrontée au problème d'optimisation combinatoire. Quant à la mise en œuvre de MPM, le problème qui se pose est le calcul des probabilités marginales des classes.

Dans le travail publié par Geman et Geman [12], il a été montré qu'en utilisant la relaxation stochastique et en modélisant l'information contextuelle spatiale contenue dans la configuration initiale (ou une image issue d'une classification ponctuelle) par un champ aléatoire de Markov, une estimation au sens de MAP pouvait être obtenue à l'aide de l'algorithme heuristique issu de la physique statistique. Ensuite, Maroquin *et al.* [25] ont utilisé le modèle de Markov et ont proposé une estimation directe des probabilités marginales pour une classification globale au sens de la MPM.

Dans le premier sous-paragraphe, nous discutons les principaux fondements des champs de Markov d'un point de vue mathématique et physique. La théorie des champs Markoviens a été inspirée de la mécanique statistique (modèle d'Ising). Dans le traitement d'images, nous utilisons les mêmes termes à savoir : énergie, potentiel, température...bien sûr le sens des mots est différent de celui en mécanique statistique. Dans le second sous-paragraphe, un modèle Markovien d'images est donné et les différentes méthodes de la classification Bayésienne globale sont abordées. Ainsi, le formalisme MRF-Bayes est détaillé et la fonction d'énergie à minimiser est déduite.

2. 1. Champs de Markov

L'utilisation des champs de Markov est devenue populaire depuis la publication des résultats de Geman et Geman [12]. Nous discutons ici les fondements de la théorie des champs Markoviens (MRF) [11], [26]-[28]. Nous définissons les MRF de la façon la plus générale sur les graphes. Soit $g = (S, \mathcal{X})$ un graphe où $S = \{s_1, s_2, \dots, s_N\}$ est l'ensemble des sommets (ou des sites) et $\mathcal{X}$ est l'ensemble des arêtes.

2. 1. 1. Voisins, système de voisinage et clique

Deux points s_i et s_j sont voisins s'il existe une arête $e_{ij} \in \mathcal{X}$ entre eux. L'ensemble des points qui sont voisins d'un site s (c'est à dire le voisinage de s) est noté V_s . Le système $V = \{V_s | s \in S\}$ est un système de voisinage pour g si $s \notin V_s$ et $s \in V_r \Leftrightarrow r \in V_s$.

Un sous-ensemble $C \subseteq S$ est une clique si chaque paire de sites dans C est voisine. Soit V note l'ensemble des cliques et $\deg(V) = \max_{C \in V} |C|$.

A chaque site du graphe, nous attribuons une étiquette l appartenant à un ensemble fini des étiquettes L . Un tel étiquetage est appelé une configuration w qui a une certaine probabilité $P(w)$. La restriction de w à un sous-ensemble $T \subset S$ est notée par w_T et $w_s \in L$ désigne l'étiquette au site s . Dans ce qui suit, nous nous intéressons aux mesures de probabilité assignées à l'ensemble de toutes les configurations possibles $\mathcal{W}$.

2. 1. 2. Définition du Champ de Markov

La variable aléatoire X est un champ de Markov par rapport au système de voisinage V si :

- pour tous les $w \in \mathcal{W} : P(X = w) > 0$,
- pour tous les $s \in S$ et $w \in \mathcal{W} : P(X_s = w_s | X_r = w_r, r \neq s) = P(X_s = w_s | X_r = w_r, r \in V_s)$.

En utilisant cette définition, nous pouvons définir une mesure de Gibbs sur $\mathcal{W}$. Soit $\mathbf{V}$ une fonction potentielle qui assigne un nombre $\mathbf{V}_c(w)$ à chaque sous configuration w_c . la fonction d'énergie sur $\mathcal{W}$ est définie par :

$$U(w) = -\sum_T \mathbf{V}_T(w) \quad (1)$$

2. 1. 3. Distribution de Gibbs

La distribution de Gibbs est une mesure de probabilité p sur $\mathcal{W}$ avec la représentation suivante :

$$p(w) = \frac{1}{Z} \exp(-U(w)), \text{ avec } Z = \sum_w \exp(-U(w)) \quad (2)$$

où Z est la constante de normalisation. Le théorème suivant fait la liaison entre les champs de Markov et la distribution de Gibbs [4], [27].

2. 1. 4. Théorème d'Hammersly-Clifford

La variable aléatoire X est un champ de Markov par rapport au système de voisinage V si et seulement si $p(w) = P(X = w)$ est une distribution de Gibbs, c'est à dire :

$$p(w) = \frac{1}{Z} \exp\left(-\sum_{c \in V} \mathbf{V}_c(w)\right) \quad (3)$$

L'intérêt du théorème d'Hammersly-Clifford est de spécifier un champ de Markov de façon simple à partir d'une formulation énergétique au lieu de caractéristiques probabilistes locales ce qui est généralement plus difficile à calculer.

2. 1. 5. Topologie du système de voisinage

En traitement d'images, S est considérée comme une grille L telle que $\forall s \in S : s = (i, j)$ et nous définissons les systèmes de voisinage homogènes d'ordre n :

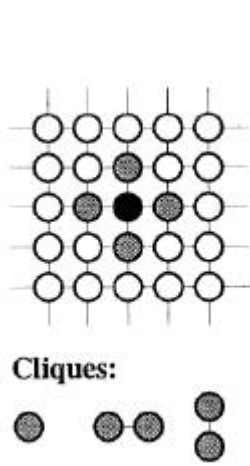
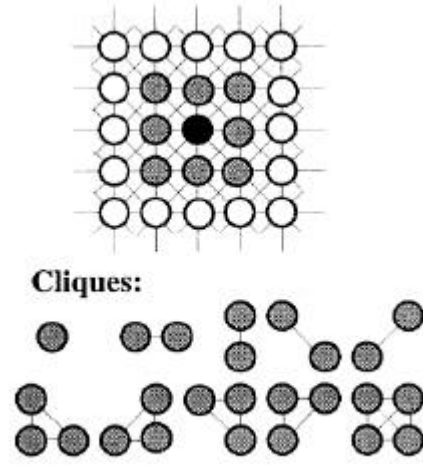
$$V^n = \{V_{(i,j)}^n : (i, j) \in L\} \quad (4)$$

$$V_{(i,j)}^n = \{(k, l) \in L : (k - i)^2 + (l - j)^2 \leq n\} \quad (5)$$

Il est clair que $V^0 \equiv S$ et pour tous les $n \geq 0 : V^n \subset V^{n+1}$. Les figures 1 et 2 montrent des systèmes de voisinage d'ordre un ($n = 1$) et d'ordre deux ($n = 2$), respectivement. Les cliques sont $\{(i, j)\}$, $\{(i, j), (i, j + 1)\}$, $\{(i, j), (i + 1, j)\}$. En pratique, les systèmes de voisinages d'ordre supérieur à deux ne sont pas utilisés car leur fonction d'énergie est trop compliquée et nécessitent un temps de calcul élevé.

2. 2. Un Modèle Markovien Général d'Images

Nous présentons ici la formulation mathématique générale d'un modèle Markovien d'images. Soit $R = \{r_1, r_2, \dots, r_K\}$ l'ensemble des sites et $Y = \{Y_r : r \in R\}$ l'ensemble des données (ou observations) sur ces sites. L'ensemble de toutes les observations possibles $y = \{y_{r1}, y_{r2}, \dots, y_{rN}\}$ est noté par F . Nous avons un autre ensemble de sites $S = \{s_1, s_2, \dots, s_N\}$, chacun de ces sites peut prendre une étiquette de $L = \{1, 2, \dots, M\}$.

**Figure 1.** Système de voisinage d'ordre un.**Figure 2.** Système de voisinage d'ordre deux.

L'espace de configuration $\mathcal{W}$ est l'ensemble de tous les étiquetages possibles $w = \{w_{s1}, w_{s2}, \dots, w_{sN}\}$, $w_s \in \mathcal{L}$. Les deux ensembles R et S ne sont pas nécessairement disjoints. Le but est de modéliser les étiquettes et les observations avec un champ aléatoire joint $(X, Y) \in \mathcal{W} \times \mathcal{F}$. Le champ $X = \{X_s\}_{s \in S}$ est appelé le champ des étiquettes et $Y = \{Y_r\}_{r \in R}$ est appelé le champ des observations.

2. 2. 1. Estimateur Bayésien

Tout d'abord, nous construisons un estimateur Bayésien de champ des étiquettes. Nous pouvons exprimer la probabilité jointe ainsi que la probabilité conditionnelle par les distributions *a priori* et *a posteriori* :

$$P_{X,Y}(w, y) = P_X(y|w)P_X(w) \quad (6)$$

$$P_X(w|y) = \frac{P_{X,Y}(w, y)}{P_Y(y)} = \frac{P_X(y|w)P_X(w)}{P_Y(y)} \quad (7)$$

$P_Y(y)$ est constant car la réalisation du champ des observations est connue :

$$P_X(w|y) \propto P_X(y|w)P_X(w) \quad (8)$$

L'estimateur est donné par la fonction de décision d suivante :

$$d : \mathcal{F} \rightarrow \mathcal{W} \Rightarrow y \mapsto d(y) = \hat{w} \quad (9)$$

Soit $C(w, d(y))$ la fonction de coût. Le risque de Bayes est donné par :

$$r(P, \mathbf{d}) = E\{C(w, \mathbf{d}(y))\} \quad (10)$$

L'estimateur doit avoir un risque de Bayes minimal tel que :

$$\hat{w} = \arg \min_{w \in W} \int_{w \in W} C(w, \tilde{w}) P_{\mathcal{Y}}(w|y) dw \quad (11)$$

Dans ce qui suit, nous présentons les trois estimateurs Bayésiens les plus connus [29].

2.2.2. Estimateur maximum *a posteriori* (MAP)

L'estimateur MAP est le plus souvent utilisé en traitement d'images. Sa fonction de coût est définie par :

$$C(w, \tilde{w}) = 1 - D_{\tilde{w}}(w) \quad (12)$$

où $D_{\tilde{w}}(w)$ est la masse de Dirac en $\tilde{w}$. Il est clair que cette fonction donne le même coût pour chaque configuration différente de $\tilde{w}$. En utilisant l'équation (11) et l'équation (12), l'estimateur MAP du champ des étiquettes est donné par :

$$\hat{w}_{MAP} = \arg \max_{w \in W} P_{\mathcal{Y}}(w|y) \quad (13)$$

Cet estimateur, pour une observation donnée, fournit les modes de la distribution *a posteriori*. Cependant, l'équation (13) pose un problème d'optimisation combinatoire et par conséquent, exige l'utilisation d'algorithmes spécifique d'optimisation.

2.2.3. Estimateur modes *a posteriori* marginales (MPM)

La fonction de coût de l'estimateur MPM est définie par :

$$C(w, \tilde{w}) = \sum_{s \in S} (1 - D_{\tilde{w}_s}(w_s)) \quad (14)$$

Cette fonction de coût est reliée au nombre de sites $s \in S$ où $w_s \neq \tilde{w}_s$. La solution de l'équation (11) est donnée par :

$$\forall s \in S : \hat{w}_s^{MPM} = \arg \max_{w_s \in L} P_{\mathcal{Y}}(w_s|y) \quad (15)$$

La fonction (15) permet de déterminer la configuration qui maximise à chaque site la distribution marginale *a posteriori* $P_{\mathcal{Y}}(\cdot|y)$.

2. 2. 4. Estimateur champ moyen “*Mean Field (MF)*”

La fonction de coût de l'estimateur MF est donnée par :

$$C(w, \tilde{w}) = \sum_{s \in S} (w_s - \tilde{w}_s)^2 \quad (16)$$

En utilisant l'équation (11) et l'équation (16), l'estimateur MF du champ des étiquettes est alors donné par :

$$\forall s \in S : \hat{w}_s^{MF} = \int_{w \in W} w_s \cdot P_{MF}(w|y) dw \quad (17)$$

L'estimateur (17) est la valeur moyenne conditionnelle de X sachant $Y = y$, c'est à dire le champ moyen de X .

2. 2. 5. La distribution *a priori*

Supposons que X est un MRF avec le système de voisinage $\tilde{V} = \{\tilde{V}_s : s \in S\}$ dont la distribution est définie par :

$$P(X = w) = \frac{1}{Z} \exp(-\tilde{U}(w)) \quad (18)$$

$$\tilde{U}(w) = \sum_{c \in \tilde{V}} \tilde{V}_c(w) \quad (19)$$

où $\tilde{U}(w)$ est la fonction d'énergie. Cette représentation utilise pour la définition de la probabilité *a priori* la distribution de Gibbs dont l'avantage est que l'on peut travailler avec les énergies potentielles sur les cliques au lieu de l'énergie globale.

2. 2. 6. Formalisme Bayésien-MRF pour la classification contextuelle d'images

Nous présentons ici le formalisme Bayésien-MRF pour la classification contextuelle d'images. Le but est de trouver une image des étiquettes (image classifiée) qui maximise la probabilité *a posteriori* $P(w|Y)$, c'est à dire l'estimateur MAP du champ des étiquettes :

$$P(w|Y) = \frac{1}{P(Y)} P(Y|w) P(w) \quad (20)$$

$P(Y)$ ne dépend pas de l'étiquetage w et nous supposons que :

$$P(Y|w) = \prod_{s \in S} P(y_s|w_s) \quad (21)$$

L'étiquetage recherché sera donné par :

$$\hat{w} = \arg \max_{w \in W} \prod_{s \in S} P(y_s | w_s) \prod_{c \in V} \exp(-V_c(w_c)) \quad (22)$$

Cette équation montre que la probabilité *a posteriori* définit également un champ Markovien.

2.2.7. La distribution *a posteriori*

Les observations sont reliées au processus des étiquettes par le modèle qui modélise la relation entre le champ des étiquettes X et le processus des observations Y . La plupart des problèmes peuvent se formaliser par la fonction suivante [30], [31] :

$$Y = Y(H(X), B) \quad (23)$$

Au niveau de chaque pixel de l'image, nous avons la fonction suivante :

$$\forall r \in R : Y_r = Y(H_r(X_{Y(r)}), B_r) \quad (24)$$

où $Y(a, b)$ est une fonction inversible en a . H_r est une fonction locale définie sur un petit sous-ensemble $Y(r)$ de S tel que $Y(r) \in S, |Y(r)| < |S|$ et $Y^{-1}(s) = \{r \in R | s \in Y(r)\}$. B est une composante aléatoire dont la distribution est telle que $P_b(.) = \prod_{r \in R} P_{B_r}(.)$. Nous avons alors la relation suivante :

$$P_{Y|X}(y|w) = \prod_{r \in R} P_{B_r}(Y^{-1}(H_r(w_{Y(r)}), y_r)) \quad (25)$$

En supposant que $P_b(.) > 0$ en chaque site $r \in R$, la distribution conditionnelle du champ des observations Y sachant X , est définie par :

$$P_{Y|X}(y|w) = \exp\left(\sum_{r \in R} -\log(P_{B_r}(Y^{-1}(H_r(w_{Y(r)}), y_r)))\right) \quad (26)$$

En combinant l'équation (23) avec l'équation (8) et l'équation (18), la distribution *a posteriori* se met sous la forme suivante :

$$P_{Y|X}(w|y) \propto \frac{1}{Z} \exp\left(\sum_{r \in R} -\log(P_{B_r}(Y^{-1}(H_r(w_{Y(r)}), y_r))) + \sum_{c \in V} \tilde{V}_c(w)\right) \quad (27)$$

On remarque que la distribution *a posteriori* est aussi une distribution de Gibbs avec le système de voisinage V le plus petit qui contient tous les cliques dans $\tilde{V}$ et les ensembles $\{Y(r), r \in R\}$:

$$\forall s \in S : V_s = \left(\bigcup_{r \in Y^{-1}(s)} Y(s) \{s\} \right) \cup \tilde{V}_s \quad (28)$$

Notons la fonction d'énergie correspondante par $U(w, y)$:

$$U(w, y) = \sum_{r \in R} -\log(P_{\tilde{Y}}(Y^{-I}(H_r(w_{Y(r)}), y_r))) + \sum_{C \in \tilde{V}} \tilde{V}_C(w) = \sum_{r \in R} V_r((w_{Y(r)}), y_r) + \sum_{C \in \tilde{V}} \tilde{V}_C(w) \quad (29)$$

Nous définissons maintenant $V_r((w_{Y(r)}), y_r)$ d'une manière précise [30], [31] :

$$V_r((w_{Y(r)}), y_r) = V_r(w_{Y(r)}) + \sum_{s \in Y(r)} V_{sr}(w_s, y_r) \quad (30)$$

Donc, elle peut se mettre sous la forme :

$$\sum_{r \in R} V_r((w_{Y(r)}), y_r) = \sum_{r \in R} V_r(w_{Y(r)}) + \sum_{r \in R} \sum_{s \in Y(r)} V_{sr}(w_s, y_r) \quad (31)$$

$$\sum_{r \in R} V_r((w_{Y(r)}), y_r) = \sum_{r \in R} V_r(w_{Y(r)}) + \sum_{s \in S} \sum_{r \in Y^{-I}(s)} V_{sr}(w_s, y_r) = \sum_{r \in R} V_r(w_{Y(r)}) + \sum_{s \in S} V_s(w_s, y_{Y^{-I}(s)}) \quad (32)$$

Finalement, nous avons la fonction d'énergie suivante :

$$U(w, y) = \sum_{s \in S} V_s(w_s, y_{Y^{-I}(s)}) + \sum_{C \in \tilde{V}} V_C(w) = U_1(w_s, y_{Y^{-I}(s)}) + U_2(w) \quad (33)$$

où les énergies potentielles $V_C(w)$ des cliques sont définies par :

$$V_C(w) = \begin{cases} \tilde{V}_C(w) & \text{si } C \in \tilde{V} \text{ et } C \notin \{Y(r), r \in R\} \\ V_r(w_{Y(r)}) & \text{si } C = Y(r) \text{ et } Y(r) \notin \tilde{V} \\ \tilde{V}_C(w) + V_r(w_{Y(r)}) & \text{si } C = Y(r) \text{ et } Y(r) \in \tilde{V} \end{cases} \quad (34)$$

Si l'on suppose que l'image observée Y en s ne dépend que du pixel s , l'équation (33) se simplifie : $Y(r)$ se réduit à s et le système de voisinage de la distribution *a posteriori* est équivalent à celui de la distribution *a priori*.

La fonction d'énergie globale donnée dans (33) est la somme des deux termes suivants :

- Terme d'attache aux données caractérisé par la fonction d'énergie $U_1(w_s, y_{Y^{-I}(s)})$. Ce terme est directement calculé à partir des observations, pour cette raison il est appelé attache aux données. Par exemple, dans le cas d'images optiques multispectrales, où les classes sont supposées Gaussiennes, l'expression du terme d'attache aux données a la forme suivante :

$$U(w, Y) = \sum_{s \in S} \left\{ (y_s - m_{ws})^T \cdot S_{ws}^{-1} (y_s - m_{ws}) + \log(|S_{ws}^{-1}|) \right\} \quad (35)$$

où m_{ws} et S_{ws} , sont le vecteur moyenne et la matrice de covariance de classe w_s , respectivement.

- Terme de régularisation caractérisé par la fonction d'énergie $U_2(w)$. Ce terme traduit les interactions locales dans un voisinage spatial. C'est à travers ce terme que la classification optimale d'images est obtenue en utilisant des algorithmes d'optimisation.

2. 2. 8. Modélisation du terme de régularisation de la fonction d'énergie

Le deuxième terme de la fonction d'énergie globale, $U_2(w)$, peut être modélisé par un modèle énergétique simple de forme :

$$U_2(w) = \sum_{C \in \mathcal{V}} V_2(w_C), \quad \text{où} \quad V_2(w_C) = V_{\{s,r\}}(w_s, w_r) = \begin{cases} -b & \text{si } w_s = w_r \\ +b & \text{si } w_s \neq w_r \end{cases} \quad (36)$$

où b est l'hyperparamètre qui contrôle le degré d'homogénéité de la régularisation au niveau des cliques d'un voisinage. Il est initialisé d'une façon empirique. Cependant, le modèle donné en (33) ne décrit pas d'une manière efficace les interactions locales entre les sites dans un voisinage. Ainsi, d'autres modèles ont été utilisés dans la littérature, on cite parmi eux le modèle de Gauss, le modèle d'Ising et le modèle de Potts.

2. 2. 8. 1. Modèle de Gauss:

Le modèle de Gauss d'énergie locale est donné par :

$$U_2(w) = \sum_{C \in \mathcal{V}} V_2(w_C), \quad \text{où} \quad V_2(w_C) = V_{\{s,r\}}(w_s, w_r) = b \cdot \sum_{r \in V_s} (w_s - w_r)^2 + \tilde{b}(w_s - m_C) \quad (37)$$

où m_C est la moyenne locale dans la clique C et $\tilde{b} > 0$ est un paramètre qui pondère la dispersion d'une étiquette par rapport à la moyenne des étiquettes locales. Le modèle de Gauss est un terme de régularisation attractif ou répulsif selon le signe de b . Il permet de synthétiser des textures très diverses.

2. 2. 8. 2. Modèle d'Ising :

Le modèle d'Ising, le plus ancien, est lié à l'étude du ferromagnétisme en physique statistique. L'ensemble des étiquettes est binaire $L = \{+1, -1\}$. En traitement d'images, le modèle d'Ising est très adapté au problème de détection de contours. La fonction d'énergie locale de ce modèle est donnée par :

$$U_2(w) = \sum_{C \in \mathcal{V}} V_2(w_C), \quad \text{où} \quad V_2(w_C) = V_{\{s,r\}}(w_s, w_r) = -J \cdot \sum_{r \in V_s} (w_s \cdot w_r) \quad (38)$$

Le paramètre J est la constante de couplage entre sites voisins. Cette énergie représente un terme d'interaction qui porte sur les cliques d'ordre deux et qui conditionne la régularité du modèle. Quand le paramètre J est positif (respectivement négatif) l'énergie diminue lorsque les sites voisins sont de même signe (respectivement de signes opposés), le potentiel est dit attractif (respectivement répulsif). Le paramètre J joue le même rôle que le paramètre b dans le modèle de Gauss et le modèle de Potts.

2. 2. 8. 3. Modèle de Potts :

L'extension du modèle d'Ising à un ensemble d'étiquettes non binaires $L = \{1, 2, \dots, M\}$ où M est le nombre de classes, est donnée par le modèle de Potts. Le voisinage est

généralement celui de 4 ou 8 connexités. Seules les cliques d'ordre deux interviennent dans l'énergie de régularisation définie par :

$$U_z(w) = \sum_{C \in \mathcal{V}} V_z(w_C), \quad \text{où } V_z(w_C) = V_{\{s,r\}}(w_s, w_r) = -b \cdot \sum_{r \in V_s} d(w_s, w_r) \quad (39)$$

où d est la fonction de Kroeneker défini par :

$$d(w_s, w_r) = \begin{cases} +1 & \text{si } w_s = w_r \\ 0 & \text{si } w_s \neq w_r \end{cases}$$

Sous cette forme, avec $b > 0$, l'énergie du modèle de Potts diminue d'autant plus que les voisins d'un pixel appartiennent à la même région.

2. 2. 8. 4. Première variante du modèle de Potts :

Une variante du modèle de Potts a été définie par Jhung *et al.* [13]. Elle s'écrit sous la forme suivante :

$$U_z(w) = \sum_{C \in \mathcal{V}} V_z(w_C), \quad \text{où } V_z(w_C) = V_{\{s,r\}}(w_s, w_r) = b \cdot \sum_{r \in V_s} (1 - d(w_s, w_r)) \quad (40)$$

Sous cette forme, avec $b > 0$, l'énergie du modèle de Potts calcule le nombre de voisins ayant une étiquette différente du site central. Donc, l'énergie diminue d'autant plus que les voisins d'un pixel appartiennent à la même région.

2. 2. 8. 5. Deuxième variante du modèle de Potts :

Une autre variante du modèle qui consiste à modifier le modèle de Potts de telle sorte que le paramètre de régularisation b soit variable proportionnellement au nombre maximal de sites appartenant à la même classe dans le voisinage considéré. L'expression d'une telle variante est la suivante :

$$U_z(w) = \sum_{C \in \mathcal{V}} V_z(w_C), \quad \text{où } V_z(w_C) = V_{\{s,r\}}(w_s, w_r) = -b \frac{q - \sum_{r \in V_s} (1 - d(w_s, w_r))}{q} \quad (41)$$

où q est le nombre de sites dans le voisinage appartenant à la même classe du site central, c'est à dire ayant pour classe w_s . La valeur de b ici est le nombre de sites appartenant à la classe la plus présente dans le voisinage. Le modèle permet d'éviter de faire des tests répétitifs pour le choix de la valeur appropriée pour b , mais le problème de la préservation des structures reste toujours posé.

2. 2. 8. 6. Choix judicieux de la valeur du paramètre b :

Tous les modèles du terme de régularisation que nous venons de citer sont en fonction du hyperparamètre b dont sa valeur est déterminant de la précision du résultat de la

classification. Le calcul de la valeur adaptative de b en fonction des cliques de voisinage de chaque site a fait l'objet de plusieurs études récentes [32]-[36]. On sait qu'une valeur élevée pour b favorise les classes étendues en détriment des classes étroites. Par contre, une valeur assez petite pour b ne permet pas d'homogénéiser les classes de façon adéquate. Dans [33] et [56], une valeur critique du hyperparamètre b , notée b_c , est donnée en fonction de nombre de classes pour lesquelles une phase de transition entre l'homogénéité des régions et la préservation des structures linéaires est détectée. La valeur de b_c est calculée selon l'expression suivante [56] :

$$b_c = \log(1 + \sqrt{M}) \quad (42)$$

où M est le nombre de classes. Dans notre cas, nous avons utilisé l'expression (39) pour délimiter la fourchette des valeurs du hyperparamètre b utilisées dans les tests pour trouver la valeur la plus approprié. Ainsi, les deux limites des valeurs de b sont calculées telles que : $b_{min} = \sqrt{M} - b_c$ et $b_{max} = \sqrt{M} + b_c$.

3. Optimisation

La méthode de classification contextuelle déduite du formalisme Bayésien-MRF donne une fonction d'énergie non-convexe (expression (30)) qui doit être minimisée pour trouver l'estimateur de champ des étiquettes (image classifiée optimale). Malheureusement, c'est un problème très dur, appelé problème combinatoire. Si nous considérons, par exemple, une image de taille 16×16 avec deux étiquettes (classes) possibles, nous devons explorer un espace de configurations de 2^{256} éléments pour trouver le minimum global de la fonction d'énergie (30) qui correspond à l'image classifiée recherchée. Donc, en pratique, il est impossible de calculer toutes les valeurs possibles de la fonction d'énergie.

Cependant, ce problème combinatoire peut être exprimé sous la forme générale d'un « problème d'optimisation ». Dans la littérature, deux sortes de problèmes d'optimisation reçoivent l'appellation, non strictement définie, de problèmes combinatoires (difficiles) :

- Certains problèmes d'optimisation combinatoire, pour lesquels on ne connaît pas d'algorithme exact et rapide.
- Certains problèmes d'optimisation combinatoire, pour lesquels on ne connaît pas d'algorithmes permettant de repérer un optimum global à coup sûr, et en un nombre fini de calculs.

Pour résoudre ces problèmes d'optimisation combinatoire, il existe des méthodes déterministes, utilisant l'approche mathématique traditionnelle telle que la descente du

gradient. Ces méthodes, qui garantissent souvent l'optimalité des solutions, s'avèrent rapidement gourmandes en temps de calcul. Dans notre cas d'application, l'utilisation de méthodes classiques de descente du gradient n'est pas possible à cause de la non-convexité de la fonction d'énergie. La recherche d'un compromis entre la qualité de la solution du problème à optimiser et le temps nécessaire pour trouver cette solution a suscité un grand intérêt auprès des chercheurs opérationnels. Ils se sont mis alors à la recherche de nouvelles méthodes approchées, ou heuristiques, qui permettent d'obtenir une solution de bonne qualité, moyennant un temps de calcul plus raisonnable.

3. 1. Méthodes d'Optimisation Heuristiques et Métaheuristiques

Le mot « heuristique » vient du grec *heurein* (découvrir) et qualifie tout ce qui sert à la découverte, à l'invention et à la recherche. Pour l'algorithmique, les heuristiques sont des méthodes qui cherchent à approcher une solution optimale; on les appelle parfois méthodes approchées. Une heuristique peut être conçue pour résoudre un type de problème donné, ou bien être conçue comme une méthode générale, qui peut être adaptée à divers problèmes d'optimisation: dans le second cas, elle est désignée sous le terme de « métaheuristique ». Les métaheuristiques sont à l'origine dédiées aux problèmes combinatoires [39].

Ainsi, un grand nombre d'heuristiques, qui produisent des solutions proches de l'optimum, ont été développées pour les problèmes d'optimisation combinatoire difficiles. Il existe une abondante littérature qui traite de ces heuristiques, étudie leur convergence et les tests d'arrêt associés [18]. L'objectif n'est pas d'obtenir un optimum absolu, mais seulement une bonne solution, et la garantie de l'inexistence d'une solution sensiblement meilleure et ceci au bout d'un temps de calcul raisonnable. Une autre caractéristique d'une heuristique est la qualité de la solution obtenue, c'est-à-dire son écart avec la solution optimale. Les heuristiques disposent en général d'un certain nombre de paramètres de contrôle, dont le choix est primordial pour la qualité de la solution à obtenir. Une heuristique est « robuste » si elle converge le plus souvent vers la même solution, et elle est d'autant plus « efficace » qu'elle donne, en un temps de calcul donné, une solution meilleure, plus proche de l'optimum.

Depuis environ 15 ans, les « métaheuristiques », principalement le recuit simulé, la méthode de recherche tabou et les algorithmes génétiques, se sont imposées comme des moyens efficaces pour la résolution de problèmes d'optimisation combinatoire. La méthode du recuit simulé repose sur une analogie avec le processus de recuit employé par les métallurgistes, les algorithmes génétiques s'inspirent de l'évolution des espèces dans la nature, et la méthode de recherche tabou exploite le concept de mémoire. Ces trois approches, totalement différentes, permettent en principe de trouver une bonne solution aux problèmes d'optimisation, moyennant un temps de calcul acceptable.

3. 2. Optimisation Globale et Optimisation Locale

Tout d'abord, et afin d'enlever toute ambiguïté dans ce qui suit, nous donnons ici les définitions précises des termes minimum local et minimum global. Le problème considéré est celui de trouver le ou un minimum global de la fonction d'énergie, donnée dans l'expression (30), et d'éviter d'être piégé dans un minimum local. Nous disons que $U(\tilde{w})$ est minimum local si nous avons :

$$\exists e > 0 \mid \forall w \in W : \|w - \tilde{w}\| < e \Rightarrow U(w) \geq U(\tilde{w}) \text{ et } \tilde{w} \in W \quad (43)$$

$U(\tilde{w})$ est un minimum global si nous avons :

$$U(\tilde{w}) = \min_{w \in W} \{U(w)\} \quad (44)$$

Ainsi, dans le reste de ce chapitre, le terme d'optimisation globale fait référence à la recherche des optima globaux de la fonction d'énergie, au sens défini ci-dessus. De ce point de vue, la méthode d'optimisation globale vise la détermination des optima globaux du problème, en évitant le piègeage dans l'un de ses optima locaux. Cette dénomination présente néanmoins une certaine ambiguïté, car on rencontre souvent dans la littérature la dénomination de « méthode locale », qui fait cette fois référence au mécanisme de recherche, lorsqu'il procède par voisins successifs. Le principe est qu'à partir d'une solution de départ, considérée temporairement comme étant la valeur minimale, on engendre par perturbations élémentaires une suite finie de voisins. Ainsi, le recuit simulé est une méthode de recherche locale (la solution testée est une voisine de la solution courante), qui sera de notre point de vue une méthode d'optimisation globale (la méthode est en principe capable de déterminer les optima globaux de la fonction d'énergie). De ce fait, et pour enlever toute ambiguïté possible, nous appelons les méthodes d'optimisation locales, celles qui permettent de déterminer un minimum local et les méthodes d'optimisation globale, celles qui s'efforcent de déterminer un optimum global.

Cependant, ces méthodes ne s'excluent pas mutuellement. Afin d'améliorer les performances d'une recherche, plusieurs auteurs [20] combinent les deux types d'algorithmes. Une recherche globale permet de bien explorer l'espace de recherche et de déterminer les zones prometteuses (susceptible de contenir un minimum global) ; cette phase est appelée « diversification » ; et une recherche locale permet d'approfondir la recherche dans chacune des zones prometteuses localisées lors de l'exploration du domaine de recherche ; cette phase est appelée « intensification ». La phase d'intensification se différencie de la phase de diversification par le réglage des paramètres de contrôle de la méthode.

Dans la littérature, les méthodes d'optimisation globale sont réparties, selon les auteurs, en différentes classes [37]. Pour Goldberg, selon les critères de robustesse, les méthodes

d'optimisation globale sont subdivisées en trois types [38]: les méthodes déterministes, les méthodes énumératives, et les méthodes stochastiques utilisant un processus aléatoire guidé. Les méthodes déterministes et les méthodes énumératives ne sont pas efficaces pour tout type de problème. Lorsqu'on veut résoudre un problème difficile, dès que la dimension du problème est grande, ces méthodes peuvent avoir des temps de calcul déraisonnables, et on a alors recours aux méthodes stochastiques.

Les méthodes stochastiques sont itératives et comportent trois éléments principaux : un mécanisme de perturbation, un critère d'acceptation et un critère d'arrêt. Ils sont appliqués à partir d'un ou plusieurs points de la fonction objectif, choisis aléatoirement. La perturbation aléatoire consiste à perturber, à la fois ou un par un, tous les éléments de la configuration courante, ou une partie seulement. Elle suit une loi de distribution particulière dans un voisinage du point courant, par exemple une loi de distribution uniforme. Le ou les nouveaux points sont générés selon une loi qui dépend des points précédents et/ou des valeurs précédentes de la fonction objectif, ou ne dépend que du point courant. Le critère d'acceptation consiste à accepter le ou les nouveaux points selon une certaine loi de probabilité, ou selon certaines conditions, ou conserver les meilleurs points pour l'étape suivante. Une recherche locale est alors effectuée à partir du ou des nouveaux points. Les critères d'arrêt peuvent être liés à la qualité du minimum obtenu. La procédure est arrêtée quand il n'y a plus d'amélioration de la solution après un certain nombre d'itérations, ou elle s'arrête quand les perturbations ne dépassent plus un certain seuil. Ils peuvent être liés aussi à des coûts de calcul ; la procédure s'arrête après un nombre maximal d'évaluations de la fonction objectif, ou un nombre d'itérations fixé, ou un certain temps de calcul fixé.

Parmi les différentes méthodes stochastiques d'optimisation globale, nous allons uniquement nous intéresser aux métaheuristiques « modernes » [18]. L'objectif est d'apporter une contribution à une plus grande diffusion des méthodes métaheuristiques, dans la résolution des problèmes combinatoires posés la classification contextuelle d'images de télédétection, où les applications pratiques sont très nombreuses. Ainsi, Les trois sous-paragraphe suivants sont consacrés aux trois métaheuristiques d'optimisation globale, à savoir : le recuit simulé, la méthode de recherche tabou et les algorithmes génétiques. L'application de ces méthodes dans la classification contextuelle d'images est alors détaillée ainsi que les procédures de leur mise en œuvre.

3. 3. Recuit Simulé

Le recuit simulé trouve ses origines dans la thermodynamique [15], [16], [40]. Cette méthode est issue d'une analogie entre le phénomène physique de refroidissement lent d'un corps en fusion, qui le conduit à un état solide, de basse énergie. Il faut abaisser

lentement la température, en marquant des paliers suffisamment longs pour que le corps atteigne l'équilibre thermodynamique à chaque palier de température et éviter toute cristallisation dans un état d'énergie non minimal (minimum local). Pour les matériaux, cette stratégie de diminution de l'énergie se manifeste par l'obtention, à la fin du processus de refroidissement, d'une structure régulière, comme dans les cristaux et l'acier.

En 1953, Metropolis *et al.* [41] ont proposé une simulation Monte-Carlo pour trouver les états d'équilibre des systèmes thermodynamiques. Dans les années 80, Kirkpatrick *et al.* [15], et Cerny [40] ont montré l'analogie entre la minimisation d'une fonction non-convexe et l'état d'équilibre des systèmes thermodynamiques. Ils ont substitué la fonction d'énergie du solide à la fonction de coût à minimiser et exécutent l'algorithme de Metropolis en utilisant une séquence de température décroissant lentement. Ils ont appelé ce nouvel algorithme recuit simulé [11], [16], [17], [31], [44]. La recherche dans ce domaine est devenue intensive et a abouti à des contributions variées importantes telle que celle du l'échantillonneur de Gibbs proposé par Geman et Geman [12].

3.3.1. Dynamique de Metropolis

L'analogie exploitée par le recuit simulé consiste à considérer une fonction d'énergie à minimiser (l'expression (30) dans notre cas), et une configuration w peut être considérée comme un état donné du système thermodynamique dont $U(w)$ est l'énergie. Le recuit simulé exploite généralement le critère défini par l'algorithme de Metropolis [41] pour l'acceptation d'une configuration obtenue par perturbation de la configuration courante.

Pour une température T donnée, à partir d'une configuration courante w , on considère une perturbation qui changerait w en h . Si cette perturbation induit une diminution de la valeur de la fonction d'énergie, $DU = U(h) - U(w) < 0$, elle est acceptée. Dans le cas contraire, si $DU = U(h) - U(w) \geq 0$, la perturbation est acceptée tout de même avec une probabilité $p = \exp(-DU/T)$. Ainsi, le pseudo-code de la règle de Metropolis peut être formulé comme suit :

```

Si  $U(h) \leq U(w)$ 
     $U(w) := U(h)$ 
     $w := h$ 
Sinon  $p := \exp(-DU/T)$ 
     $r := \text{un nombre aléatoire uniforme dans } [0,1]$ 
    Si  $r \leq p$ 
         $U(w) := U(h)$ 
         $w := h$ 
    Fin si

```

3.3.2. Modèle Mathématique du Recuit Simulé

Le modèle mathématique du recuit simulé a été étudié en détails par Aarts et van Laarhoven [16]. Nous présentons ici ce modèle adapté au problème de classification contextuelle d'images. Soit $w, h, \dots$ les configurations d'un problème d'optimisation combinatoire (ils correspondent aux états d'un système thermodynamique et aux images classifiées dans notre cas). Soit $U(w)$ le coût (fonction d'énergie donnée dans l'expression (30)) de la configuration w . Les éléments d'une configuration sont indexés par $S = \{s_1, s_2, \dots, s_N\}$, où N représente dans notre cas le nombre de pixels de l'image à classifier, et l'espace d'états possibles est noté par $L\{1, 2, \dots, M\}$, où M est le nombre de classes dans notre cas. L'ensemble de toutes les configurations possibles est noté par W . Puisque $\forall s \in S : w_s \in L$, on a alors $W = L^N$.

Le recuit simulé génère une séquence de configurations (images classifiées) qui constituent une chaîne de Markov avec une probabilité de transition $P_{w,h}(k-1, k)$, où k caractérise un équilibre thermodynamique qui correspond à un palier de température. Soit $X(k)$ l'état (image classifiée dans notre cas) atteint après la $k^{ième}$ transition. La probabilité de cet événement est donnée par :

$$P(X(k) = w) = \sum_z P(X(k-1) = z) P_{z,w}(k-1, k) \quad k = 1, 2, \dots \quad (45)$$

Si la probabilité de transition $P_{w,h}(k-1, k)$ ne dépend pas de k , la chaîne correspondante est homogène, sinon elle est inhomogène. Les probabilités de transition dépendent aussi de la température T . Donc, si T est constant, la chaîne est homogène et la matrice de transition $P = P(T)$ peut être écrite comme suit :

$$P_{w,h}(T) = \begin{cases} G_{w,h}(T) A_{w,h}(T) & \forall h \neq w \\ 1 - \sum_z G_{w,z}(T) A_{w,z}(T) & h = w \end{cases} \quad (46)$$

où $G_{w,h}(T)$ est la probabilité de générer h à partir de w et $A_{w,h}(T)$ est la probabilité d'acceptation de la configuration h . Il est clair que $P = P(T)$ est une matrice stochastique :

$$\forall w : \sum_z P_{w,z}(T) = 1 \quad (47)$$

La température T influe sur la probabilité d'accepter une configuration plus mauvaise. A une température élevée, la probabilité d'acceptation d'une perturbation quelconque tend vers un : presque tous les changements sont acceptés. L'algorithme équivaut alors à une marche aléatoire dans l'espace des configurations. Cette température est diminuée lentement au fur à mesure du déroulement de l'algorithme pour simuler le processus de

refroidissement des matériaux, et sa diminution est suffisamment lente pour que l'équilibre thermodynamique soit maintenu.

3.3.3. Paramètres de Contrôle du Recuit Simulé

L'efficacité du recuit simulé dépend fortement du choix de ses paramètres de contrôle, dont le réglage reste très empirique. Ces paramètres de contrôle sont les suivants :

- Valeur initiale de la température.
- Fonction de décroissance de la température.
- Critère de changement de palier de température.
- Critères d'arrêt (ou température finale).

La température initiale T_0 doit être choisie telle que toutes les transitions peuvent être acceptées avec une probabilité non nulle. Il est très difficile de trouver une telle valeur initiale car elle est reliée au minimum et maximum d'énergie [12]. Pour le calcul de la température de départ, plusieurs méthodes ont été proposées. Une des méthodes est basée sur l'observation de la variation moyenne de la fonction d'énergie. A partir d'une configuration initiale w_0 on génère, par perturbations aléatoires, un certain nombre de configurations $\tilde{w}_0$ (environ 50 à 100) telles que $U(\tilde{w}_0) > U(w_0)$, et on calcule la variation moyenne $\langle DU \rangle_{init}$. Une température initiale T_0 est calculée de façon à accepter au départ une certaine proportion p_{init} de perturbations dégradant la fonction d'énergie. Pour une température initiale moyenne, la valeur de p_{init} est de 0.5. La valeur de T_0 est déduite de la formule suivante :

$$p_{init} = \exp\left(\frac{-\langle DU \rangle_{init}}{T_0}\right) \quad (48)$$

Le rôle de la température T au cours du processus de recuit simulé est très important. Une forte décroissance de température risque de piéger l'algorithme dans un minimum local, alors qu'une faible décroissance au début du processus entraîne une convergence très lente de l'algorithme. Or, il est connu que le recuit simulé converge avec une probabilité de un vers un optimum global si la loi de décroissance de la température est moins rapide que $Cst/\log(k)$ pour une certaine constante Cst qui est indépendante de k [12], [16], [45]. La loi logarithmique de décroissance de la température, qui assure la convergence théorique du recuit simulé, peut prendre la forme suivante :

$$T_k = \frac{q}{\log(1+k)} \quad (49)$$

où k est le nombre de paliers de température effectués, et q une constante positive. En pratique, la loi théorique est approchée par une loi exponentielle car elle est trop lente. Ainsi, on trouve dans la littérature la proposition de la loi de la décroissance géométrique suivante [15], [42], [43] :

$$T_{k+1} = \alpha T_k, \text{ avec } 0 < \alpha < 1 \quad (50)$$

Pour le changement de palier de température, on peut simplement spécifier un nombre de perturbations, acceptées ou non, au bout duquel la température est abaissée.

Pour la température finale (ou le critère d'arrêt), évidemment, $\lim_{k \rightarrow \infty} T_k = 0$ peut être seulement approximée par un nombre fini de valeurs T_k . Donc, nous avons besoin d'un critère d'arrêt. Le critère le plus souvent utilisé en pratique est d'arrêter l'exécution après un certain nombre d'itérations, ou bien si DU est inférieure à un certain seuil.

3. 3. 4. Implémentation du Recuit Simulé

Dans notre application du modèle mathématique du recuit simulé, pour la classification contextuelle d'images, la probabilité de générer h à partir de w , $G_{w,h}(T)$, est une distribution uniforme sur les configurations h qui sont différents de w en un seul élément. La probabilité d'acceptation, $A(T)$, est donnée par le critère de Metropolis, tel que :

$$A_{w,h}(T) = \min\{1, \exp[-(U(h) - U(w))/T]\} \quad (51)$$

Le pseudo-code de l'algorithme de recuit simulé qui nous implémenté est le suivant :

```

w := configuration aléatoire.
U(w) := valeur de la fonction d'énergie.
Umin := U(w)
wmin := w
T := initialise r température (expression (45)).
Répéter
  Répéter
    - Construire une perturbation h à partir la configuration courante w
      telle que h est différente de w en un seul élément (pixel).
    - Calculer l'énergie de la configuration perturbée.
    - Appliquer la règle de Metropolis.
  Jusqu'à ce que l'équilibre thermodynamique atteint.
  T := décroître température (expression (46)).
Jusqu'à ce que les conditions d'arrêt soient satisfaites.

```

Cet algorithme est connu sous le nom de recuit homogène, car il est décrit par une séquence des chaînes de Markov homogènes. Si la température décroît après chaque

transition, l'algorithme est décrit par une chaîne de Markov inhomogène et il est appelé recuit inhomogène. Un tel algorithme est obtenu en supprimant la condition d'obtention de l'équilibre thermodynamique avant la décroissance de la température. Dans la pratique, c'est ce type de recuit qui est le plus souvent utilisé.

La matrice d'acceptation donnée dans (48) peut être remplacée par une autre matrice d'acceptation plus élaborée proposée par Geman et Geman [12]. Dans l'algorithme proposé [12], la matrice de génération a été remplacée par une autre matrice de génération de forme :

$$G_{w,h} = \begin{cases} 1 & \text{si } h \in N_w \\ 0 & \text{sinon} \end{cases} \quad (52)$$

où N_w est l'ensemble de configurations qui ne diffèrent de la configuration courante w que par un élément au minimum. Un échantillonnage de N_w est effectué tel que la nouvelle configuration h soit acceptée avec une probabilité de :

$$\frac{\exp(-U(h))}{\sum_{z \in N_w} \exp(-U(z))} \quad (53)$$

Cette règle jointe à un recuit inhomogène est devenue l'algorithme le plus populaire appelé l'échantillonneur de Gibbs.

3. 4. Recherche Tabou

La recherche tabou est une métaheuristique originalement développée par Glover [46] et indépendamment par Hansen [47], sous l'appellation de "*steepest ascent mildest descent*". Elle est basée sur des idées simples, mais elle est néanmoins très efficace. Cette méthode combine une procédure de recherche locale avec un certain nombre de règles et de mécanismes permettant à celle-ci de surmonter l'obstacle des optima locaux, tout en évitant de cycler [18], [39]. Elle a été appliquée avec succès pour résoudre de nombreux problèmes difficiles d'optimisation combinatoire : problèmes de routage de véhicule [48], problèmes d'affectation quadratique [49], problèmes d'ordonnancement [50], problèmes de coloration de graphes [51], etc.... Ici, nous considérons l'application de la méthode recherche tabou pour résoudre le problème d'optimisation combinatoire posé par la classification contextuelle d'images [52].

3. 4. 1. Principe de Base

Dans une première phase, la méthode de recherche tabou peut être vue comme une généralisation des méthodes d'amélioration locales. En effet, en partant d'une configuration quelconque w appartenant à l'ensemble de configurations W , on se déplace

vers une configuration h située dans le voisinage de w . Donc l'algorithme explore itérativement l'espace de configurations W .

Afin de choisir le meilleur voisin h dans le voisinage de w , l'algorithme évalue la fonction d'énergie en chaque configuration voisine, et retient le voisin qui améliore la valeur de la fonction d'énergie, ou au pire celui qui la dégrade le moins.

L'originalité de la méthode de recherche tabou, par rapport aux méthodes locales, qui s'arrêtent dès qu'il n'y a plus configuration voisine permettant d'améliorer la valeur de la fonction d'énergie, réside dans le fait que l'on retient la meilleure configuration voisine, même si celle-ci est plus mauvaise que la configuration d'où l'on vient. Ce critère autorisant les dégradations de la fonction d'énergie évite à l'algorithme d'être piégé dans un minimum local. Mais il induit un risque de cyclage. En effet, lorsque l'algorithme a quitté un minimum quelconque par acceptation de la dégradation de la fonction d'énergie, il peut revenir sur ses pas, à l'itération suivante.

Pour régler ce problème, l'algorithme a besoin d'une mémoire pour conserver pendant un moment la trace des dernières meilleures configurations déjà visitées. Ces configurations sont déclarées tabou, d'où le nom de la méthode. Elles sont stockées dans une liste de longueur L donnée, appelée liste tabou. Une nouvelle configuration n'est acceptée que si elle n'appartient pas à cette liste tabou. Ce critère d'acceptation d'une nouvelle configuration évite le cyclage de l'algorithme, durant la visite d'un nombre de configurations au moins égal à la longueur de la liste tabou, et il dirige l'exploration de la méthode vers des régions du domaine de configurations non encore visitées. Le pseudo-code de l'algorithme tabou classique peut être formulé comme suit :

```

w := configuration aléatoire.
U(w) := valeur de la fonction d'énergie.
Umin := U(w)
wmin := w
Tabou := liste de configurations de longueur L.
Tabou := Vide.
Répéter
  - Générer Nv configurations telle que chaque configuration h ∈ voisinage de w et
    1-) {w, h} ∉ Tabou;
    2-) U(h) = min1 ≤ i ≤ Nv U(hi).
  - Ajouter ({w, h}, Tabou).
  - w := h
  - Si U(w) < Umin
    Umin := U(w)
    wmin := w
  - Fin Si
Jusqu'à ce que les conditions d'arrêt soient satisfaites.

```

La liste tabou est généralement gérée comme une liste « circulaire » : on élimine à chaque itération la configuration tabou la plus ancienne, en la remplaçant par la nouvelle configuration retenue. Mais le codage d'une telle liste est encombrant, car il faudrait garder en mémoire tous les éléments qui définissent une configuration. Pour pallier à cette contrainte, on remplace la liste tabou de configurations interdites par une liste de « perturbations interdites », en interdisant la perturbation inverse d'une perturbation faite récemment.

3. 4. 2. Critère d'Aspiration

Le remplacement de la liste tabou des configurations visitées par la liste des perturbations élémentaires $\{w, h\}$ conduit non seulement à l'interdiction de revenir vers des configurations précédentes (on évite le cyclage court), mais aussi vers un ensemble de configurations dont plusieurs peuvent ne pas avoir été visitées jusqu'ici. Il est donc primordial de corriger ce défaut et de trouver un moyen de lever l'interdiction de l'acceptation d'une perturbation élémentaire $\{w, h\}$ déjà effectuée (donc appartenant à la liste tabou), sous un certain critère, appelé critère d'aspiration. Cette correction permet aussi de revenir à une configuration déjà visitée et de redémarrer la recherche dans une autre direction. Cette idée est développée dans [53] , [54] et [20].

Le critère d'aspiration le plus simple et le plus couramment utilisé consiste à tester si la configuration produite de statut tabou présente un coût inférieur à celui de la meilleure configuration trouvée jusqu'à présent. Si cette situation se produit, le statut tabou de la configuration est levé. Ce critère est évidemment très sévère, il ne devrait pas être vérifié très souvent, donc il apporte peu de changements à la méthode. D'autres critères d'aspiration plus complexes peuvent être envisagés. L'inconvénient de recourir trop souvent à l'aspiration est qu'elle peut détruire, dans une certaine mesure, la protection offerte par la liste tabou vis-à-vis du cyclage.

Notons que, dans le cas d'une liste tabou de configurations, le concept de critère d'aspiration n'est pas intéressant. Toute annulation du statut tabou d'une configuration se trouvant dans la liste tabou pourrait conduire l'algorithme au cyclage.

3. 4. 3. Intensification

L'intensification consiste à approfondir la recherche dans certaines régions du domaine, identifiées comme susceptibles de contenir un optimum global. Cette intensification est appliquée périodiquement, et pour une durée limitée. Pour mieux intensifier la recherche dans une zone bien localisée, plusieurs stratégies sont proposées dans la littérature.

La plus simple consiste à retourner à l'une des meilleures configurations trouvées jusqu'à présent, puis de reprendre la recherche à partir de cette configuration, en réduisant la longueur de la liste tabou pour un nombre limité d'itérations. Dans ce cas, on adapte la procédure de recherche tabou, en élargissant le voisinage de la configuration courante, tout en diminuant le pas des perturbations. On peut aussi remplacer simplement l'heuristique tabou par une autre méthode plus puissante, ou mieux adaptée, pour une recherche locale.

3. 4. 4. Diversification

La diversification permet à l'algorithme de bien explorer l'espace des configurations, et d'éviter que le processus de recherche ne soit trop localisé et laisse de grandes régions du domaine totalement inexplorées. La plus simple des stratégies de diversification consiste à interrompre périodiquement l'acheminement normal de la procédure tabou, et à la faire redémarrer à partir d'une autre configuration, choisie aléatoirement, ou « intelligemment ». Une autre méthode consiste à biaiser la fonction d'énergie, en introduisant un terme qui pénalise les perturbations effectuées fréquemment, afin de favoriser des perturbations nouvelles. Ce type de stratégie de diversification peut être utilisé de façon continue, sans interrompre la procédure de recherche tabou.

En résumé, nous dirons que la diversification et l'intensification sont des concepts complémentaires, qui enrichissent la méthode de recherche tabou et la rendent plus robuste et plus efficace.

3. 4. 5. Implémentation de la Méthode Recherche Tabou

Tout d'abord, nous devons spécifier deux paramètres de la méthode de recherche tabou appliquée pour résoudre le problème d'optimisation combinatoire posé par la classification contextuelle d'images. Le premier paramètre est la longueur de la liste tabou. Dans notre cas, ce paramètre peut être fixé, au début du processus itératif de régularisation, au nombre total de classes existant dans l'image. Ce nombre est considéré comme étant la taille maximale, notée TM , de la liste tabou. Puis, au fur et mesure ce nombre changera d'une manière adaptative en fonction de nombre d'étiquettes des pixels du voisinage de chaque pixel de l'image. Ce voisinage peut être de taille 4 ou 8 selon la topologie du voisinage choisie par l'utilisateur. Le deuxième paramètre est le voisinage de la configuration considérée, noté N_v . Notons ici qu'il ne s'agit pas du voisinage du pixel utilisé dans la modélisation Markovienne, mais du voisinage des étiquettes des pixels voisins. Autrement dit, la taille de ce voisinage est dépendante du nombre d'étiquettes des pixels voisins. Ainsi, durant le processus itératif, la taille et le contenu de ce voisinage sont changeables. Le pseudo-code de la méthode recherche tabou que nous avons implémenté est alors le suivant :

```

w := configuration aléatoire.
U(w) := valeur de la fonction d'énergie.
Umin := U(w)
wmin := w
Tabou := liste de configurations de taille maximale TM.
Tabou := Vide.
Répéter
- Générer Nv configurations (h) telle que h ∈ voisinage de w,
  et elle est différente de w en un seul élément (pixel).
- Calculer les énergies des configurations h, U(h).
- Réordonner les configurations h de manière à ce que
  les énergies correspondantes apparaissent par valeur décroissantes.
- Si h n'est pas tabou, ou si h est tabou mais U(h) > Umin
                                     (critère d'aspiration)
                                     w := h
                                     htabou := h,
                                     Aller à l'étape suivante.

- Sinon w := h
      htabou := h, où U(h) est la plus grande des valeurs U(h), ... U(hNv)
      h est non tabou.
- Ajouter htabou en queue de liste tabou et poser Tabou := Tabou + 1.
- Si Tabou dépasse TM
      Ôter l'élément situé en tête de liste
      Tabou := TM.

- Fin Si
- Si Umin < U(w)
      wmin := w
      Umin := U(w).
- Fin Si
Jusqu'à ce que les conditions d'arrêt soient satisfaites.

```

Les conditions d'arrêt sont réalisées lorsqu'un nombre prédéfini d'itérations de déroulement des séquences est atteint, ou bien que l'énergie minimale, U_{min} , atteigne un seuil prédéfini.

3. 5. Algorithmes Génétiques

Les principes fondamentaux de ces algorithmes ont été exposés par Holland [57]. Ces algorithmes s'inspirent du fonctionnement de l'évolution naturelle, notamment la sélection de Darwin, et la procréation selon les règles de Mendel.

La sélection naturelle, que Darwin appelle l'élément « propulseur » de l'évolution, favorise les individus d'une population qui sont le mieux adaptés à un environnement. La sélection est suivie de la procréation, réalisée à l'aide de croisements et de mutations au niveau du patrimoine génétique des individus (ou « génotype »), constitué d'un ensemble

de gènes. Ainsi deux individus « parents », qui se croisent, transmettent une partie de leur patrimoine génétique à leurs descendants. Le génotype de l'enfant fait que celui-ci est plus ou moins bien adapté à l'environnement. S'il est bien adapté, il a une plus grande chance de procréer dans la génération future. Au fur et à mesure des générations, on sélectionne les individus les mieux adaptés, et l'augmentation du nombre d'individus bien adaptés fait évoluer la population entière.

Dans ce qui suit, nous allons présenter d'abord le principe de base des algorithmes génétiques. Puis, la présentation des différents codages est donnée. Ensuite nous décrirons les opérateurs génétiques, avant de rappeler quelques résultats théoriques. L'application des algorithmes génétiques pour résoudre le problème combinatoire de la classification contextuelle est abordée dans le dernier sous-paragraphe.

3. 5. 1. Principe de Base

Dans les algorithmes génétiques, on essaie de simuler le processus d'évolution d'une population d'individus (chromosomes). Chaque individu est constitué d'un ensemble d'éléments appelés caractéristiques ou gènes pouvant prendre plusieurs valeurs appartenant à un alphabet non nécessairement numérique. Le degré d'adaptation d'un individu à l'environnement est exprimé par la valeur de la fonction d'énergie $U(w)$, où w est la configuration que l'individu représente. On dit qu'un individu est d'autant mieux adapté à son environnement, que l'énergie de la configuration qu'il représente est plus faible. Le but est donc de trouver la meilleure combinaison des éléments des individus afin d'atteindre le maximum d'adéquations. Pour atteindre cet objectif l'algorithme travaille en parallèle sur une population d'individus candidats (chromosomes).

On part d'une population de N configurations du problème représentées par des individus. Cette population choisie aléatoirement est appelée population parent. Au sein de cette population, intervient alors la sélection au hasard d'un ou deux parents, qui produisent une nouvelle configuration, à travers les opérateurs génétiques, tels que le croisement et la mutation. A chaque itération, appelée génération, est créée une nouvelle population d'individus. Cette nouvelle génération est constituée généralement par des individus mieux adaptés à l'environnement. Ces individus sont créés en utilisant des parties des meilleurs éléments de la population précédente ainsi que des parties innovatrices. En itérant ce processus, on produit une population plus riche en individus mieux adaptés. Au fur et à mesure des générations, les individus vont tendre vers l'optimum de la fonction d'énergie. Ainsi, l'algorithme génétique de base peut être résumé en trois phases distinctes :

- Production de la population d'individus la mieux adaptée pour contribuer à la reproduction de la génération suivante ; elle peut être mise en œuvre sous plusieurs formes algorithmiques.
- Phase de reproduction, qui exploite essentiellement les opérateurs de croisement et de mutation.
- Stratégie de remplacement des populations parent et enfant par la génération suivante. Elle pourra être mise en œuvre sous plusieurs formes.

Ainsi, le pseudo-code du principe d'un algorithme génétique de base peut être formulé comme suit :

```

Pop, Popn et Qualité : Tableaux de taille N
Pop := initialiser population
 $U(w) := \min_{1 \leq i \leq N} (U(w_i))$ 
 $U_{min} := U(w)$ 
 $w_{min} := w$ 
Répéter
  - Qualité (évaluer la population Pop).
  - Répéter (phase de reproduction génétique)
    - Sélection
    - Croissement
    - Mutation
  - Jusqu' à ce que Popn soit remplie
  - Pop := sélectionner nouvelle population (Pop, Popn).
  -  $U(w) := \min_{1 \leq i \leq N} (U(w_i))$ 
  - Si  $U(w) < U_{min}$ 
    -  $U_{min} := U(w)$ 
    -  $w_{min} := w$ 
  - Fin Si
Jusqu' à ce que les conditions d'arrêt soient satisfaites.

```

3. 5. 2. Codage

Les algorithmes génétiques nécessitent le codage des paramètres à optimiser sous forme d'un ou plusieurs individus, et opèrent sur le code plutôt que sur les paramètres eux-mêmes. Le choix du codage a un effet sur les performances d'un algorithme génétique. D'une part, on doit choisir un codage simple dans le sens où le temps de décodage est minimal, d'autre part, il doit être valide dans le sens où les résultats obtenus doivent correspondre aux contraintes du problème. Plusieurs codes d'informations sont utilisés. Les plus fréquemment utilisés sont le code binaire naturel et le code binaire de Gray. Plus récemment le codage « réel » a fait son apparition.

3. 5. 2. 1. Codage binaire :

Ce codage consiste, pour un individu donné, à concaténer toutes ses variables codées en binaire. Par exemple, un gène est un entier long (32 bits). Un chromosome est un tableau de gènes (Figure 4). Un individu est un tableau de chromosomes. La population est un tableau d'individus. On aboutit à une structure constituée de cinq niveaux d'organisation (Figure 3). Ce codage binaire présente plusieurs avantages : alphabet minimum, facilité de mise en place d'opérateurs génétiques et existence de résultats théoriques [38]. Néanmoins, ce codage présente trois inconvénients majeurs :

- Les performances de l'algorithme sont diminuées lorsque la longueur de la chaîne augmente.
- Deux nombres décimaux voisins (exemple : 7 et 8) peuvent être très éloignés dans le codage binaire naturel (1000 et 0111) : falaise de Hamming. Ce problème peut être réglé en remplaçant le code binaire naturel par le code binaire de Gray.
- La dissymétrie entre le 0 (plus fréquent) et le 1.

3. 5. 2. 2. Codage réel :

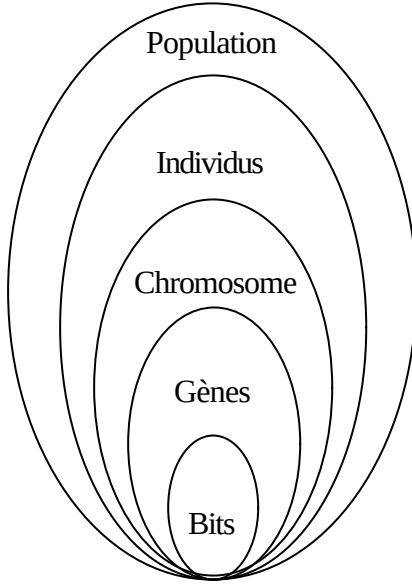
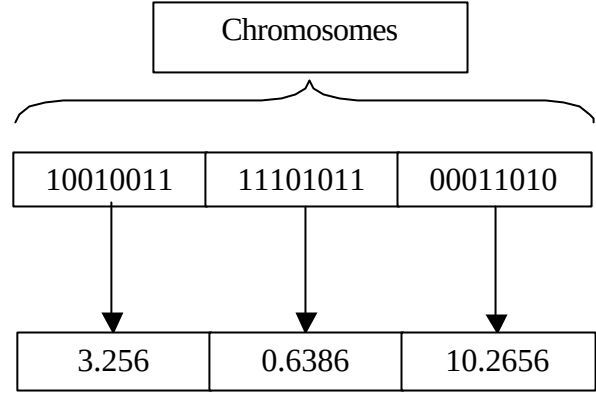
Ce codage consiste simplement à la concaténation des éléments d'un individu. Exemple un individu (25, 31, 8) est codé 25| 31| 8 Ce codage présente des avantages majeurs. Il est plus précis que le codage binaire et l'espace de recherche est le même que l'espace du problème. Il a le mérite d'être simple à utiliser, l'évaluation de la fonction d'énergie est plus rapide. Le codage réel évite de faire le transcoding du binaire naturel ou de Gray vers les réels à chaque évaluation. Néanmoins, il possède deux inconvénients, son alphabet est infini, et il a besoin d'opérateurs appropriés.

3. 5. 3. Population Initiale

La population initiale représente un ensemble d'individus de taille N , chaque individu (chromosomes) est une configuration potentielle du problème. La population initiale est générée aléatoirement. Cependant certaines méthodes garantissent le démarrage avec des configurations assez bonnes, obtenues à l'aide des algorithmes approchés.

3. 5. 4. Fonction d'Evaluation

C'est l'une des concepts les plus importants et les plus compliqués dans les algorithmes génétiques. Elle est appelée aussi fonction d'adéquation. Elle évalue chaque configuration en utilisant certains critères qui vont définir son aptitude à donner la bonne configuration au problème. Dans notre cas, la fonction d'adéquation est la fonction d'énergie. Ainsi, on associe à chaque individu un poids qui va définir l'existence ou l'élimination de cet individu dans la prochaine génération.

**Figure 3.** Organisation d'un algorithme génétique.**Figure 4.** Codage binaire.

3. 5. 5. Calcul de la Qualité

On détermine la population d'individus les mieux adaptés. On calcule la qualité de chaque individu de la population pour déterminer sa probabilité de sélection. Plus la qualité est élevée, mieux l'individu est adapté. On obtient cette qualité à partir de la fonction d'énergie par une perturbation telle que :

$$Qualité(w) = U_{max} - U(w) \quad (54)$$

U_{max} devra être choisi de manière à ce que la valeur de $Qualité(w)$ reste toujours positive.

3. 5. 6. Opérateurs de Reproduction

La phase de reproduction exploite principalement deux opérateurs : le croisement et la mutation. Elle comporte aussi l'opération de sélection et l'opération de production de la génération suivante, à partir des populations parent et enfant.

3. 5. 6. 1. Sélection :

La sélection consiste à choisir les paires d'individus qui vont participer à la reproduction de la population future. La fonction de sélection calcule une probabilité de sélection pour chaque individu, en fonction de sa qualité et de la qualité de tous les autres individus dans la population. Ainsi, l'individu i est sélectionné avec une probabilité p_{si} , telle que :

$$p_{si} = \frac{f_i}{\sum_{i=1}^N f_i} \quad (55)$$

où f_i représente la qualité d'adéquation de l'individu i , N est la taille de population. L'opérateur de sélection est appliqué n fois, c'est à dire autant de fois qu'il y a d'individus dans la population, car la taille des générations est constante. Pour réaliser la sélection, il suffit de faire tourner une roue un nombre de fois égale à la taille de la population [60]. Les individus les plus adaptés ont plus de chance d'être sélectionnés, ils peuvent même être sélectionnés plusieurs fois.

3. 5. 6. 2. Croisement :

Le principal opérateur agissant sur la population de parents est le croisement "Crossover en anglais", qui est appliqué avec une certaine probabilité, appelée taux de croisement P_c (typiquement proche de l'unité). En pratique, seulement une partie de la population, sélectionnée aléatoirement, participe à cette opération. Le nombre attendu d'individus qui subissent le croisement est donné par : $N.P_c$.

Le croisement consiste à choisir deux individus représentés par leurs chaînes de gènes, tirés au hasard dans la population courante, et à définir aléatoirement un ou plusieurs points de croisement. Les nouvelles chaînes sont alors créées en inter-changeant les différentes parties de chaque chaîne. Les deux descendants ainsi obtenus sont appelés des enfants. Cet opérateur permet de bien explorer le domaine de variation des individus, et de diriger la recherche vers des régions intéressantes de l'espace d'étude en utilisant la connaissance déjà présente dans la population courante. Il y a trois types de croisements standards :

- Croisement uni-point, où on choisit au hasard un site de croisement correspondant à un nombre d'individus entre 1 et l , l étant la longueur du chromosome [60]. Soit k l'indice de site, le 1^{er} enfant sera formé à partir des k premiers gènes du 1^{er} parent et $(l-k)$ derniers gènes du second parent, le 2^{ème} enfant sera formé à partir des k premiers gènes du 2^{ème} parent avec les $(l-k)$ derniers gènes du 1^{er} parent. Un exemple de ce type de croisement est donné dans la figure 5.
- Croisement bi-point qui consiste à choisir aléatoirement deux points sur la longueur de chacun des deux parents (les deux points sont les mêmes pour les deux parents). On obtient donc trois parties pour chaque parent (un segment tête, un segment centre et un segment queue). Le 1^{er} enfant contient : les (1^{ère} et dernière) parties du 1^{er} parent, la partie du milieu de l'enfant contiendra la partie centre du 2^{ème} parent. La construction du 2^{ème} enfant suit le même principe en inversant les parents. La figure 6 montre un exemple typique de croisement bi-point.
- Croisement uniforme, où il y a deux possibilités de croisements. Dans la première possibilité, le 1^{er} enfant sera produit de la manière suivante : on prend le 1^{er} gène du 1^{er} parent concaténé au second gène du 2^{ème} parent concaténé au 3^{ème} gène du 1^{er} parent et

ainsi de suite jusqu'à ce que la longueur de l'enfant soit égale à celle des parents. Le 2^{ème} enfant sera construit de la même manière en partant du 2^{ème} parent. La figure 7 montre un exemple de cette possibilité de croisement uniforme. Dans la deuxième possibilité, chaque gène est créé en copiant le gène du 1^{er} parent ou du 2^{ème} parent, dans la même position selon un masque de croisement généré aléatoirement. Le 1^{er} enfant est conçu de la manière suivante : un « 1 » dans le masque signifie que le gène du 1^{er} parent est copié dans le 1^{er} enfant et un « 0 » signifie que le gène du 2^{ème} parent est copié dans le 1^{er} enfant. Le 2^{ème} enfant est conçu de la manière suivante : un « 1 » dans le masque signifie que le gène du 2^{ème} parent est copié dans le 2^{ème} enfant et un « 0 » signifie que le gène du 1^{er} parent est copié dans le 2^{ème} enfant. La figure 8 illustre un exemple de cette possibilité de croisement uniforme.

3. 5. 6. 3. Mutation :

L'opération de mutation protège les algorithmes génétiques des pertes prématurées d'informations pertinentes. Elle permet d'introduire une certaine information dans la population, qui a pu être perdue lors de l'opération de croisement. Ainsi elle participe au maintien de la diversité, utile à une bonne exploration du domaine de recherche. L'opérateur de mutation s'applique avec une certaine probabilité, appelée taux de mutation P_m , typiquement compris entre 0.05 et 0.10, sur la valeur d'un ou plusieurs gènes d'un individu. Ce faible taux de mutation permet de dire que la mutation est considérée comme un mécanisme d'adaptation secondaire pour les algorithmes génétiques. Lorsqu'un gène est muté, sa valeur est remplacée par une autre appartenant au même alphabet. On estime le nombre de gènes qui vont subir l'opération de mutation par :

$$N.P_m.B \quad (56)$$

où B est le nombre de bits du chromosome. Dans le codage binaire, la mutation consiste à changer un bit 1 par le bit 0 et réciproquement, pour chaque bit de la chaîne, avec la probabilité P_m . La figure 9 représente l'opération de mutation pour un individu représenté par une chaîne binaire codée sur 8 bits.

Une autre façon de faire consiste à choisir un individu à muter avec la probabilité P_m , et à changer un bit de cet individu choisi au hasard. Cette seconde méthode est plus intéressante, car la probabilité de mutation P_m est indépendante de la longueur de codage des variables des individus.

L'algorithme génétique fait évoluer une population. La sélection réduit la diversité de la population, tandis que les opérateurs génétiques, croisement et mutation, augmentent cette diversité.

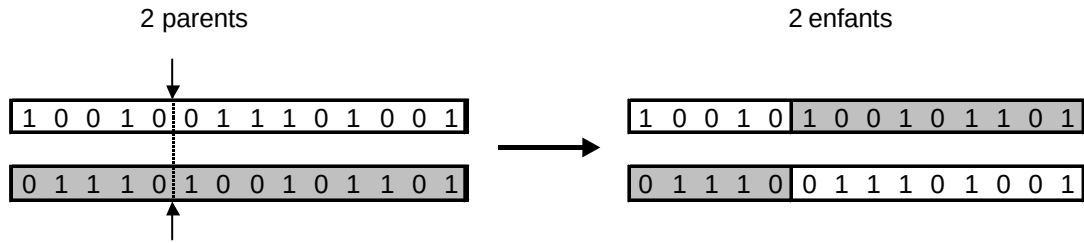


Figure 5. Représentation schématique du croisement uni-point.

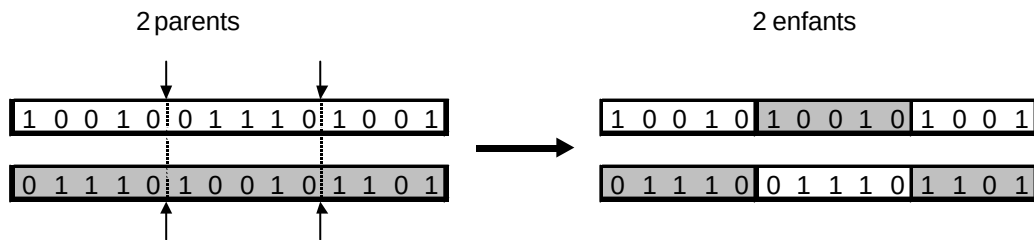


Figure 6. Représentation schématique du croisement bi-point.

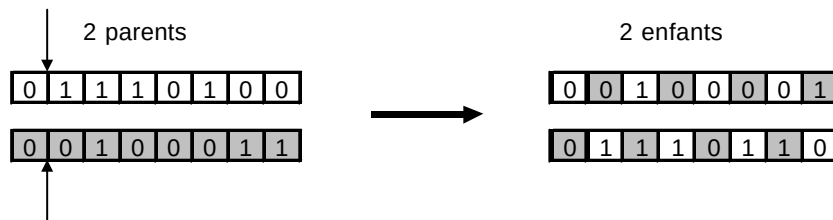


Figure 7. Représentation schématique du croisement uniforme (première possibilité).

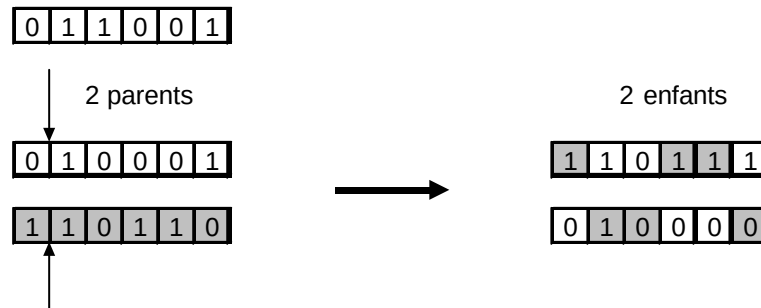


Figure 8. Représentation schématique du croisement uniforme (deuxième possibilité).

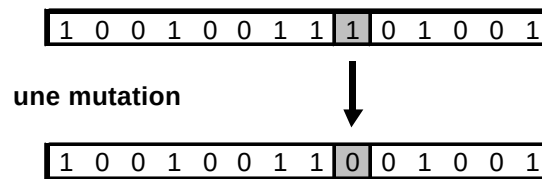


Figure 9. Représentation schématique d'une mutation dans un chromosome.

3. 5. 7. Quelques Résultats Théoriques

Le choix d'un codage adéquat est un élément critique dans l'efficacité d'un algorithme génétique. Il existe deux types de difficultés dans le choix d'un codage.

3. 5. 7. 1. Principe des alphabets minimaux :

Un codage doit pouvoir être adapté au problème de façon à limiter au mieux la taille de l'espace de recherche, et aussi de façon que les nouveaux individus engendrés par les opérateurs de recherche soient significatifs le plus souvent possible [58] [23].

3. 5. 7. 2. Principe de pertinence des schémas :

Une théorie qui décrit l'évolution des configurations partielles, appelées schémas, dans une population d'une génération à l'autre est introduite par Holland [57]. Par exemple, la chaîne *1101* appartient aux régions *11***, **10**, *11*1*, etc.: les * désignent indifféremment un *1* ou un *0*. L'efficacité d'un algorithme génétique dépend des schémas associés aux bonnes régions (i.e. contenant un nombre élevé de bonnes configurations). Le principal résultat de cette théorie est que d'une génération à une autre, la répétition des croisements accroît les bons motifs dans la population de manière exponentielle [38], à condition que ces bons motifs soient courts. Le facteur de croissance dépend beaucoup de paramètres de l'algorithme tels que le taux de croisement et le taux de mutation. La multiplication de ces bons motifs appelés blocs de construction accroît évidemment les chances de trouver la configuration optimale recherchée.

Les algorithmes génétiques ont relativement peu de fondements théoriques. Il n'existe aucune garantie que la méthode trouve la configuration optimale, bien que des travaux récents donnent des preuves de convergence [59]. Ces travaux tiennent peu compte de l'influence de l'opérateur de croisement, qui est réduit parfois à sa plus simple forme, afin de simplifier l'analyse de l'algorithme uniquement en fonction de la mutation et de la sélection. Il est difficile de faire l'analyse d'un algorithme génétique dans sa globalité, mais en revanche on peut s'intéresser à l'un de ses opérateurs séparément.

3. 5. 8. Programmation Génétique

La programmation génétique est une autre méthode basée sur la théorie de la sélection naturelle proposée pour la première fois par John Koza [61]. Koza s'est inspiré des concepts des algorithmes génétiques de John Holland. Pour cela, certains disent que la programmation génétique est une branche des algorithmes génétiques. Elle a apporté quelques avantages en permettant la variation des tailles des chromosomes, et en changeant les techniques de croisement. Par exemple, un croisement avec un seul père, en programmation génétique, peut donner deux fils différents, or ce n'est pas le cas dans les algorithmes génétiques. La programmation génétique consiste à générer des programmes informatiques d'une manière aléatoire. Elle est constituée de quatre étapes essentielles :

- Générer une population initiale.
- Exécuter chaque programme et lui assigner une valeur qui indique son aptitude à résoudre le problème posé.
- Créer une nouvelle population des programmes en utilisant les opérations génétiques (croisement, reproduction, mutation).
- Répéter les étapes 2 et 3 jusqu'à atteindre l'optimum.

3. 5. 8. 1. Génotype et phénotype :

Dans un programme génétique, les propriétés génétiques des individus sont contenues dans des génotypes qui sont des chaînes des codes des fonctions de base du problème. L'ensemble de ces génotypes forme la population qui va être le sujet des traitements dans la programmation génétique. L'exécution du génotype, i.e. l'exécution de la chaîne des codes de fonction de base, génère un certain résultat qui va définir un état de l'individu. Cet état qui définit un comportement ou une morphologie au autre est nommé phénotype.

3. 5. 8. 2. Fonctions de base :

Pour résoudre un problème par la programmation génétique, on doit lui déterminer d'abord l'ensemble des fonctions nécessaires. Ces fonctions doivent former au moins un résultat pour que le programme converge. Ces fonctions sont appelées fonctions de base. Le mot base ici a presque la même signification que celui qu'il a dans l'algèbre linéaire car tous les programmes seront créés à partir de ces fonctions qui seront codées pour faciliter la manipulation des individus lors de l'exécution du génotype par la fonction d'évaluation ou pendant le déroulement d'une opération génétique.

3. 5. 8. 3. Déroulement d'un programme génétique :

La programmation génétique démarre avec une population initiale d'individus qui ont des critères primaires. Les chromosomes (génotypes) sont générés avec une combinaison aléatoire des fonctions de base. L'ensemble des chromosomes de la population initiale doit former des programmes corrects syntaxiquement et sémantiquement. Une grammaire de génération de ces chromosomes est alors définie. La taille de la population initiale ainsi

que celle de ses individus seront des critères d'exécution. Un changement dans leurs valeurs peut évoquer une augmentation ou une diminution dans le nombre de générations et le temps d'exécution nécessaire pour atteindre l'objectif.

3. 5. 8. 3. a- Fonction d'évaluation :

La fonction d'évaluation a pour rôle d'exécuter les programmes contenus dans les génotypes (chromosomes) des individus. On peut dire que la fonction d'évaluation est un mini compilateur pour le langage qui a comme alphabet l'ensemble des fonctions de base et comme grammaire la grammaire de génération des individus. Ce mini compilateur ne s'intéresse pas aux erreurs lexico-syntaxiques car elles sont traitées par la fonction de génération des individus ainsi que les opérateurs génétiques (croisement, mutation). Un autre rôle de la fonction d'évaluation comme indique son nom, est d'évaluer les individus de la population. L'évaluation se fait selon des critères bien déterminés pour chaque problème. L'ensemble de ces critères doit bien définir l'objectif de l'évolution, qui va définir la condition d'arrêt du programme génétique.

Le nombre de critères d'évaluation peut rendre la conception de la fonction d'évaluation très compliquée. En effet, la fonction d'évaluation doit calculer une certaine fonction de ses critères qui donne pour chaque individu une valeur (qui sera son poids primaire). Après la fin d'exécution de chaque programme, la fonction d'évaluation affecte à l'individu un poids primaire p_i . Et après la fin d'exécution de tous les programmes, la fonction d'évaluation calcule la somme des poids des individus de la population pour affecter à chaque individu un poids final p tel que:

$$p = \frac{p_i}{\sum_{i=1}^N p_i} \quad (57)$$

Le poids de l'expression (57) est considéré comme la probabilité de sélection. Cette probabilité détermine les individus qui vont rester dans la prochaine génération pour subir les opérations génétiques.

3. 5. 8. 3. b- Reproduction :

La reproduction est tout simplement la sélection des bons individus pour la prochaine génération. Cela se fait à partir de la probabilité de sélection. Le nombre ou le pourcentage des individus sélectionnés dépend de la politique choisie.

3. 5. 8. 3. c- Croisement :

Le croisement s'applique sur deux individus choisis selon une politique précise dans la population des individus déjà sélectionnés. Il existe deux types de croisement:

- Croisement externe, où chacun des chromosomes des deux parents choisis est coupé en deux parties, et on fait la concaténation de la première partie de chacun d'eux à la

deuxième de l'autre pour avoir deux nouveaux chromosomes qui seront insérés dans la nouvelle génération. Pour introduire les fils, il faut concaténer la première partie du premier parent par la deuxième partie de l'autre, et la même chose pour le deuxième parent.

- Croisement interne, où chacun des chromosomes des deux parents choisis est coupé en trois parties au lieu de deux. Puis, on permute les deux parties du milieu des deux chromosomes.

Le problème principal dans les deux types de croisement est la validité des files obtenues, car ils doivent être corrects sémantiquement et syntaxiquement. Pour cela, nous déterminons une certaine norme de compatibilité entre l'alphabet de notre programme génétique [61].

La politique choisie, pour le croisement est l'un des critères qui peut influencer sur le temps d'exécution et le nombre de générations nécessaires pour la convergence du programme génétique. Généralement, nous ne pouvons pas connaître la politique adéquate pour un problème *a priori*, car les performances d'une politique changent d'un problème à un autre, et donc elles ne seront pas connues avant la fin d'exécution de notre programme génétique. Cependant, il existe plusieurs politiques de croisement dans la programmation génétique, parmi elles : Meilleur/ Meilleur, Meilleur/ Meilleur de tous/ Pire. Ces politiques ont généré deux genres de croisement : croisement avec deux parents différents et croisement avec le même parent.

3. 5. 8. 3. d- Mutation :

La mutation s'applique après le croisement. Elle permute la valeur d'une ou de plusieurs gènes du chromosome avec une certaine probabilité. Muter un gène c'est le remplacer par un autre appartenant au même alphabet en respectant la règle de compatibilité citée auparavant. L'opération de mutation et l'opération de croisement sont complémentaires permettant ainsi de rendre le programme génétique converge vers la solution du problème.

3. 5. 9. Classification Contextuelle Utilisant la Programmation Génétique

Tout d'abord, pour rendre le lecteur plus familier avec l'application de la programmation génétique pour résoudre le problème combinatoire posé par la classification contextuelle d'images, nous donnons les significations des terminologies utilisées par les algorithmes génétiques dans notre application.

Dans cette application de la programmation génétique, nous démarrons à partir d'un individu qui représente une image classifiée ponctuellement (la configuration initiale). La population initiale est constituée de N images, représentant la même scène observée, générées aléatoirement à partir de l'image classifiée ponctuellement (configuration

initiale). Soient $w_1, w_2, \dots, w_N$ les N images ainsi générées. Chaque image individu w_i est une configuration qui ne diffère d'une autre configuration w_j qu'à un seul pixel à la fois. L'évaluation d'une population consiste à calculer la fonction d'adéquation (l'énergie globale donnée dans l'expression (33)) pour chaque image individu i (configuration w_i). Dans la sélection, nous choisissons les images individus (configurations) les mieux adaptés (ayant des énergies globales minimales) pour générer la prochaine population. Le croisement consiste à permuter, avec une probabilité de croisement P_c , une ligne au hasard d'une configuration w_i avec une autre ligne au hasard d'une autre configuration w_j de la même population. La mutation consiste à changer, avec une probabilité de mutation P_m , plusieurs pixels dans toutes les configurations de la population. Le nombre total de pixels à changer est donné par l'expression (56). Pour le codage, nous avons opté pour le codage décimal (dans notre cas il s'agit du niveau de gris) où chaque chromosome représente l'étiquette de la classe à laquelle appartient le pixel.

Nous pouvons donc résumer l'algorithme de base de la programmation génétique appliquée dans la classification contextuelle d'images pas les étapes essentielles suivantes :

- (1) Fixer la taille de la population N .
- (2) Générer les N configurations (images individus $w_1, w_2, \dots, w_N$) constituant la population.
- (3) Fixer la valeur de la probabilité de croisement P_c (soit $P_c = 70\%$).
- (4) Fixer la valeur de probabilité de mutation P_m (soit $P_m = 20\%$).
- (5) Calculer les énergies des individus $U(w_1), U(w_2), \dots, U(w_N)$, puis faire évoluer la population.
- (6) Appliquer les opérations génétiques (sélection, croisement et mutation).
- (7) Répéter les étapes (5) et (6) un nombre prédéfini de fois jusqu'à ce qu'une configuration w_i ayant un degré d'adéquation, énergie $U(w_i)$, optimisé soit généré. w_i est l'image classifiée recherchée.

Le croisement que nous avons utilisé est le croisement interne (avec une probabilité de croisement $P_c = 70\%$). La politique appliquée est «*meilleur de tous*» i.e. le croisement se fait entre la meilleure configuration (image individu) et le reste de $(N-1)$ configurations de la population. La mutation est une opération assez importante vu son rôle dans notre cas. En fait, le croisement ne rapporte pas de nouveaux changements à la population, tandis que la mutation change des pixels et apporte de nouveaux changements à la population avec une probabilité de mutation $P_m = 20\%$, i. e. 20% des pixels de la configuration vont être changés.

Dans la mise en œuvre de la programmation génétique, nous avons introduit la notion de température afin d'arrêter le processus itératif et définir le critère de Stop. La loi

géométrique pour la décroissance de la température (expression (46)) est utilisée. Nous donnons dans ce qui suit le pseudo-code de l'algorithme de la programmation génétique que nous avons implémenté.

```

- Fixer la valeur de la température maximale  $T_{max}$ 
- Fixer la valeur de la température minimale  $T_{min}$ 
- Construire une population de  $N$  configurations à partir de la configuration initiale
  de telle sorte que chaque configuration générée ne diffère de l' autre
  que d' un seul pixel à la fois.
- Calculer l' énergie globale de chaque configuration de la population (expression (33)).
- Déterminer la meilleure configuration, i.e. celle qui a la plus petite énergie globale.
- Sauvegarder l' état de la meilleure configuration trouvée.
- Tant que ( $T_{max} < T_{min}$ )
  faire { - Sélection( );
          - Croisement ( );
          - Mutation( );
          - Evaluation( );
          - Test( );
          - Décroître la température (expression (46)). }

```

Avant de donner les significations ainsi que les pseudo-codes des fonctions, *Sélection()*, *Croisement()*, *Mutation()*, *Evaluation()* et *Test()*, nous donnons quelques notations utilisées dans l'implémentation de ces fonctions. Soit U^{wi} la valeur de l'énergie globale de la configuration w_i . Soit p_i la probabilité de sélection de la configuration w_i calculée selon l'expression (55), avec f_i est égale à la valeur de l'énergie globale de la configuration U^{wi} . La probabilité d'adaptation cumulative de la configuration w_i , notée pac_i , est calculée en fonction de la probabilité de sélection p_i , telle que :

$$pac_i = \begin{cases} pac_i = p_i \\ \sum_{j=1}^i p_i \end{cases} \quad (58)$$

La fonction *Sélection()* utilise la probabilité d'adaptation cumulative pour sélectionner la configuration convenable. Le pseudo-code d'une telle fonction est le suivant :

```

- Tirage aléatoire d'un nombre "n" entre [0,1]
- Pour  $i := 1 \rightarrow N$ 
  Faire { - Si  $n < pac_i$ 
          Sélectionner la configuration 1
        - Sinon Pour  $j := 1 \rightarrow N$ 
          Faire { - Si  $n \geq pac_j$  et  $n < pac_{j+1}$ 
                  Sélectionner la configuration (j + 1)
          - Fin Si } }

```

Le pseudo-code de la fonction *Croisement()* est le suivant :

- Tirage aléatoire d'un nombre "n" entre $[0,1]$
- Si $n < P$
 - Positionner sur une ligne "k" au hasard de la première configuration.
 - Faire une permutation de la ligne "k" de la première configuration avec la ligne "k" de la deuxième configuration.
- Fin Si.

Le pseudo-code de la fonction *Mutation()* est le suivant (les variables *Col.* et *Lig.* désignent le nombre de colonnes et le nombre de lignes, respectivement, de l'image) :

```

Pour i := 1 → N
  Faire { Pour j := 1 → Col.
    Faire { Pour k := 1 → Lig.
      Faire { - Tirage aléatoire d'un nombre "n" entre  $[0,1]$ 
        - Si  $n < P_m$ 
          - Tirage aléatoire d'une étiquette "m".
          - Pixel(i, j) := m.
        - Fin Si. }}}
    - Fin Si. }}
```

La fonction *Test()* consiste à faire le test suivant : Si la meilleure configuration de la nouvelle population est meilleure que la meilleure configuration de la population précédente, alors on sauvegarde la meilleure configuration de la nouvelle population ; sinon on remplace la mauvaise configuration de la population courante par la meilleure de la population précédente.

4. Résultats Expérimentaux

Le but de ce paragraphe est d'évaluer les performances des méthodes d'optimisation décrites dans le paragraphe précédent, en particulier sur les problèmes de classification contextuelle des images de télédétection en utilisant le formalisme Bayésien-MRF décrit dans le paragraphe 2.

Les tests ont été réalisés sur une image SAR fournie par le système AirSAR de la NASA/JPL. Nous remercions Pr S. B. Serpico et Dr F. Melgani, de l'Université de Gênes en Italie, d'avoir eu la bonté de nous fournir les images SAR du système aéroporté AirSAR. Le système a fourni neuf images SAR en trois bandes fréquentielles (bande C, bande L, et bande P) et en trois polarisations chacune (HH, HV, VV). Nous avons choisi l'image de la bande L polarisation HV pour réaliser nos expériences (image de la figure 10

après une amélioration du contraste), car elle est la moins bruitée par le speckle. L'image, acquise en juillet 1989, représente une scène d'étude de 250x350 pixels d'une région agricole située près de Feltwell en Royaume Uni (tous droits réservés à *Hunting Technical Services Ltd. (UK)*). Les champs agricoles sont représentés par cinq classes de la couverture terrestre correspondant, respectivement, à cinq types de récoltes : carottes, pommes de terre, sol nu, chaume et betteraves à sucre.

La figure 12 représente le résultat de la classification ponctuelle de l'image SAR de la figure 10. Ce résultat a été obtenu en utilisant la méthode de la classification non-supervisée floue proposée dans [62]. L'image de la figure 12 est la configuration initiale à optimiser en utilisant les trois méthodes d'optimisation : recuit simulé, recherche tabou et programmation génétique. Les images résultats obtenus avec ces trois méthodes d'optimisation sont données dans les figures 13, 14 et 15, respectivement.

Pour les trois méthodes d'optimisation, nous avons utilisé la première variante du modèle de Potts, donnée dans l'expression (40), avec 4 ou 8 connexions (figures 1 et 2). La valeur du paramètre de régularisation, b , est choisie de telle sorte qu'elle soit à l'intérieur de l'intervalle $[\sqrt{M} - b_c, \sqrt{M} + b]$ où la valeur de b_c est donnée dans l'expression (42). La température initiale, l'énergie initiale, l'énergie finale (ou le minimum atteint par la méthode), et le temps d'exécution des trois méthodes sont rassemblés dans le tableau 1.

En interprétant les images résultats finaux de la classification contextuelle (figures 13, 14, et 15), nous constatons que les méthodes d'optimisation : recuit simulé, recherche tabou et programmation génétique ont rendu l'image de la configuration initiale (figure 12) plus homogène et plus claire. Ceci se caractérise par le changement du nombre de pixels des différentes classes avant et après le processus d'optimisation (figure 11) ce qui reflète l'immigration des pixels mal classifiés dans les premières configurations d'une classe à une autre. Bien que ces méthodes rencontrent certaines difficultés dans la séparation du bruit (ou pixel mal classifié) et du pixel contour, l'aperçu général de l'image résultat de la classification contextuelle est plus net et plus parlant sur l'état de la surface de la scène imagée que celui de l'image de la configuration initiale. Ces résultats font ressortir des détails qui sont difficiles à discriminer dans l'image de la figure 12.

En général, les résultats des trois méthodes d'optimisation sont acceptables. Cependant, à titre de comparaison entre ces trois méthodes d'optimisation, nous constatons que le résultat de la méthode recherche tabou est le meilleur par rapport à ceux des méthodes de recuit simulé et programmation génétique en termes de netteté, délimitation des bords des classes et temps de calcul. Les résultats du recuit simulé et programmation génétique se ressemblent, mais leurs temps d'exécution sont différents. La programmation génétique est la méthode la plus coûteuse en temps de calcul, suivie par le recuit simulé (voir tableau 1).

Nous justifions la qualité de la méthode de la programmation génétique par le fait que la température initiale choisie est relativement petite par rapport à ce qu'il faut pour aboutir à un meilleur résultat de cette méthode [61]. Quoiqu'il en soit, l'image obtenue soit plus nette et précise, si nous augmentons la valeur de la température initiale, ceci se fait au détriment du temps de calcul qui devient prohibitif. Les mêmes remarques peuvent être reformulées pour la méthode recuit simulé qui converge théoriquement vers l'optimum global en temps infini [16].

5. Conclusion

Dans ce chapitre, nous avons discuté les deux étapes principales d'une classification contextuelle d'images de télédétection. En premier lieu, la modélisation Markovienne d'images est abordée et le formalisme Bayésien-MRF est déduit. Ceci consiste à utiliser les champs aléatoires de Markov et l'estimation Bayésienne, en particulier l'estimation par le Maximum A Posteriori (MAP). Cet estimateur, pour une observation donnée, fournit les modes de la distribution *a posteriori*. L'avantage d'un tel formalisme est que l'information *a priori* peut être caractérisée localement par les potentiels des cliques.

En second lieu, la méthode de classification contextuelle, déduite du formalisme Bayésien-MRF, donne une fonction d'énergie non-convexe qui doit être minimisée pour trouver l'image classifiée optimale. C'est un problème très dur, appelé problème d'optimisation combinatoire qui exige l'utilisation d'algorithmes spécifiques d'optimisation. Parmi les différentes méthodes d'optimisation globale, nous nous sommes uniquement intéressés aux métaheuristiques « modernes ». L'objectif étant d'apporter une contribution à une plus grande diffusion des méthodes métaheuristiques, dans la résolution des problèmes combinatoires posés par la classification contextuelle d'images de télédétection. Ainsi, trois métaheuristiques d'optimisation globale, à savoir : le recuit simulé, la méthode de recherche tabou et les algorithmes génétiques ont été développées. L'application de ces méthodes dans la classification contextuelle d'images a été alors détaillée ainsi que les procédures de leur mise en œuvre. Les résultats obtenus sont très satisfaisants.

Références

- [1] P. H. Swain and S. M. Davis, *Remote Sensing: The Quantitative Approach*. Mc Graw-Hill, 1978.
- [2] S. Chitroub, *Implémentation d'un processus d'analyse et de classification d'images satellitaires. Application à la caractérisation des états de surface*. Thèse de Magister, Spécialité Traitement d'Images de Télédétection, Institut d'Electronique, USTHB, Alger, 03/94-M/EL, May 1994.

- [3] J. A. Richards, *Remote Sensing Digital Image Analysis. An Introduction*. 2nd Edition, Springer-Verlag, Berlin, 1993.
- [4] J. Besag, "Spatial interaction and the statistical analysis of lattice system (with discussion)", *Journal of Royal Statistics Society*, B-36, pp. 192-236, 1974.
- [5] S. Chitroub, B. Sansal, "Maximum likelihood classification of polarimetric SAR images with an adaptive a priori probabilities", In *Proceedings of International Conference on Electronic Engineering (IEEA)*, 7-9 December, 1997, Batna University, Algeria, pp. 170-176.
- [6] S. Chitroub, R. Khedam, H. Belhadj et B. Sansal, "Champs aléatoires de Markov et recuit simulé pour la classification contextuelle optimisée d'image multispectrale optique", In *Proceedings of Mediterranean Conference on Electronics and Automatic (MCEA)*, September, 19-21 1998, Marrakech, Maroc, pp. 27-32.
- [7] S. Chitroub, R. Khedam, H. Belhadj, and B. Sansal, "Compound deterministic pseudo annealing-Markov random field model for contextual classification of remotely sensed imagery", *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 21-25, September 1998, Barcelona, Spain, Vol. 3500, pp. 435-446.
- [8] S. Chitroub, H. Belhadj, R. Khedam et B. Sansal, "Classification contextuelle d'images utilisant les matrices de co-occurrences de niveaux de gris (MCNG)", In *Proceedings of Conférence Maghrébine sur l'Automatique, l'Electrotechnique et l'Electronique Industriel (COMAEI)*, Béjaï a, du 7 au 8 Decembre, 1998, Algérie, pp. 105-108.
- [9] S. Chitroub, H. Belhadj, R. Khedam et B. Sansal, "Classification contextuelle d'image multispectrale utilisant la modelisation Markovienne et l'optimisation déterministe (ICM)", In *Proceedings of International Meeting on Computer & Electronic Systems*, du 17 au 18 mai 1999, Sidi Bel Abbes, Algérie, pp. 17-23.
- [10] S. Chitroub, A. Houacine, H. Belhadj, R. Khedam, and B. Sansal, "Bayesian-Markov random field formulation for contextual classification of remotely sensed imagery", In *Proceedings of Conférence Magéibine en Génie Electrique (CMGE)*, du 04 au 06 décembre 1999, Constantine, Algérie, pp. 320-324.
- [11] R. Chellappa and A. K. Jain (Editors), *Markov Random Fields: Theory and Application*. Academic Press, 1993.
- [12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of image", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 6, pp. 721-741, 1984.
- [13] Y. Jhung and P. H. Swain, "Bayesian contextual classification based on modified M-estimates and Markov random fields", *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 34, No. 1, pp. 67-74, January 1996.
- [14] S. Chitroub A. Houacine, A. Allaoua, M. D. Aroua, and B. Sansal, "A New deterministic annealing for image contextual classification," In *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP)*, June 5-9, 2000, Istanbul, Turkey, pp.2243-2246.
- [15] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing", *Science* 220, pp. 671-680, 1983.
- [16] P. J. Van Laarhoven and E. H. L. Aarts, *Simulated Annealing: Theory and Applications*. Dordrecht: D. Reidel Publisher Company, Netherlands, 1987.
- [17] R. Azencott, *Parallel Simulated Annealing: An Overview of basic Techniques*. In R. Azencott, editor, *Parallel Simulated Annealing: Parallelization Techniques*, pp. 37-46, John Willy & Sons, 1992.
- [18] C.R. Reeves (Ed.), *Modern Heuristic Techniques for Combinatorial Problems*. Advanced Topics in Computer Science, Chapter 4, McGraw - Hill, 1995.
- [19] R. Chelouah and P. Siarry, "Adaptation of several combinatorial metaheuristics for global optimization of functions with continuous variables" (submitted in 1999).
- [20] F. Glover and M. Laguna, *Tabu Search, Modern Heuristic Techniques for Combinatorial Problems*. University of Colorado, Boulder, USA, 1992.
- [21] R. Chelouah and P. Siarry, "Tabu Search applied to Global Optimization", *European Journal of Operational Research*, Vol. 123, pp. 256-270, 2000.

-
- [22] M. Krueger, *Méthode d'analyse d'algorithmes d'optimisation stochastiques à l'aide d'algorithmes génétiques*. Thèse de Doctorat en Informatique et Réseaux, Ecole nationale supérieure des télécommunications, ENST-Paris, France, 1990.
 - [23] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*,. Springer-Verlag Heidelberg, 1996.
 - [24] R. Chelouah and P. Siarry, "Continuous Genetic Algorithm : An algorithm for the global optimization of multi-minima functions", *Journal of Heuristics*, Vol. 6, pp. 191-213, 2000.
 - [25] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solution of ill-posed problems in computational vision", *Journal of the American Statistical Association*, Vol. 82, pp. 76-89, 1987.
 - [26] P. L. Dobruschin, "The Description of a random Field by means of conditional probabilities and construction of its regularity", *Theory of Probability and its Applications*, Vol. XIII, No. 2, pp. 197-224, 1968.
 - [27] J. Moussouris, "Gibbs and Markov random system with constraints", *Journal of Statistical Physics*, Vol. 10, No. 1, pp. 11-33, January 1974.
 - [28] Y. Rosanov, *Markov Random Fields*. Springer Verlag, 1982.
 - [29] J. L. Marroquin, *Probabilistic Solution of Inverse Problems*. Ph. D Thesis, Massachusetts Institute of Technology, Artificial Intelligence Laboratory, USA, 1985.
 - [30] P. Pérez, *Champs Markoviens et analyse Multirésolution de l'Image: Application à l'Analyse du Mouvement*. Thèse de Doctorat, Université de Rennes I, Rennes, France, 1993.
 - [31] Z. Kato, *Modélisation Markoviennes Multirésolutions en Vision par Ordinateur. Application à la Segmentation d'Images SPOT*. Thèse de Doctorat, Spécialité Sciences de l'Ingénieur, Université de Nice Sophia Antipolis, Nice, France, 1994.
 - [32] O. Pony, X. Descombes, et J. Zerubia, "Classification d'images satellitaires hyperspectrales en zones rural et périurbaine", *Rapport de recherche, Institut National de Recherche en Informatique et en Automatique*, No. 4008, Sophia Antipolis, France.
 - [33] R. D. Morris and X. Descombes, "The Ising/ Potts model is not well suited to segmentation tasks", In *Proceedings of IEEE Digital Signal Processing Workshop*, September 1-4, 1996, Norway.
 - [34] C. F. Borges, "On the estimation of Markov random field parameter", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 3, pp. 216-224, 1999.
 - [35] X. Descombes, R. D. Morris, J. Zerubia, M. Berthod, "Estimating Gaussian Markov random field parameters in a non-stationary framework: Application to remote sensing imagery", *IEEE Transactions on Image Processing*, Vol. 8, No. 4, pp. 490-503, 1999.
 - [36] X. Descombes, R. D. Morris, J. Zerubia, M. Berthod, "Estimation of Markov random field prior parameters using Markov chain Monte Carlo maximum likelihood", *IEEE Transactions on Image Processing*, Vol. 8, No. 7, pp. 934-963, 1999.
 - [37] A. Törn and A. Zilinskas, *Global optimization*. G. Goos, J. Hartmanis (Eds), Springer Verlag, No 350, 1989.
 - [38] D.E Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
 - [39] I. Charon, A. Germa et O. Hudry, *Méthodes d'Optimisation Combinatoire*. Collection Pédagogique de Télécommunication, Edition MASSON, 1996.
 - [40] V. Cerny, "Thermodynamical approach to the traveling salesman problem : an efficient simulation algorithm", *Journal of Optimization Theory and Applications*, Vol. 45, No. 1, pp. 41-51, January 1985.
 - [41] N. Metropolis, A.R. Rosenbluth, M.N. Rosenbluth, A. Teller, E. Teller, "Equation of state calculations by fast computing machines", *The Journal of Chemical Physics*, Vol. 21, No. 6, pp. 1087-1092, 1953.
 - [42] P. Siarry, *La méthode du recuit simulé : application à la conception de circuits électroniques*. Thèse de Doctorat de l'Université Pierre et Marie Curie, Paris 6, novembre 1986.
 - [43] P. Siarry, *La méthode du recuit simulé en électronique : adaptation et accélération. Comparaison avec d'autres méthodes d'optimisation. Application dans d'autres domaines*. Rapport d'habilitation à diriger les recherches en sciences physiques, Université de Paris Sud, Centre d'Orsay, février 1994.
-

-
- [44] J. Besag, "On the statistical analysis of dirty pictures", *Journal of Royal Statistics Society*, B-36, pp. 192-236, 1986.
 - [45] B. Hajek, "A Tutorial survey of theory and applications of simulated annealing", In *Proceedings of 24th Conference on Decision and Control*, Lauderdale, Finland, December 1985, pp. 755-760.
 - [46] F. Glover, "Future Paths for Integer Programming and Links to Artificial Intelligence", *Computers and Operations Research*, Vol. 13, No. 5, pp. 533-549, 1986.
 - [47] P. Hansen, "The Steepest Ascent Mildest Descent Heuristic for Combinatorial Programming", *Communication at the Congress on Numerical Methods in Combinatorial Programming*, Capri, Italy, 1986.
 - [48] M. Gendreau, A. Hertz & G. Laporte, "A Tabu Search Algorithm for the Vehicule Routing Problem", *Management Science*, No. 40, pp. 1276-1290, 1994.
 - [49] J. Skorin-Kapov, "Tabu search applied to the quadratic assignment problem", *ORSA Journal on Computing*, Vol. 2, pp. 3-45, 1990.
 - [50] M. Widmer & A. Hertz, "A New Method for the Flow Sequencing Problem", *European Journal of Operations Research*, No. 41, pp. 186-193, 1989.
 - [51] A. Hertz & D. de Werra, "Using Tabu Search Techniques for Graph Coloring", *Computing*, No. 29, pp. 345-351, 1987.
 - [52] S. Chitroub, A. Houacine and B. Sansal, "Tabu search-based relaxation method for image contextual classification", In *Proceedings of The First IEEE International Symposium on Signal Processing and Information Technology*, December 28-30, 2001, Cairo, Egypt, pp. 351-355.
 - [54] F. Glover, "Tabu search", *CAAI Report 88-3*, Center for Applied Artificial Intelligence, Graduate School of Business, Box 419, University of Colorado, Boulder, Colorado, 80309-0419, June 1988.
 - [55] F. Glover, E. Taillard and D. de Werra, *A User's Guide to Tabu Search*. University of Colorado and EPFL Lausanne, 1991.
 - [56] H. O. Georgii, *Gibbs Measures and Phase Transitions*. De Gruyter, Berlin, 1988.
 - [57] J.H. Holland, *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, Internal report, 1975.
 - [58] K.A. De Jong, *An Analysis of the Behavior of a Class of Genetic Adaptive Systems*. University of Michigan, Ann Arbor, MI, Ph.D Thesis, 1975.
 - [59] D.E Goldberg and K. Deb, A "Comparative Analysis of Selection Schemes Used in Genetic Algorithms", *Foundation of Genetic Algorithms*, San Mateo, California, USA, 1991, pp. 69-93.
 - [60] Jean Michel Renders, *Algorithmes génétiques et réseaux de neurones*. Edition Berti, 1994.
 - [61] J. R. Koza, *Genetic Programming*. MIT Press, Cambridge, MA, USA, 1992.
 - [62] S. Chitroub, A. Houacine, and B. Sansal, "Robust optimal fuzzy clustering algorithm applicable to multispectral and polarimetric synthetic aperture radar images", *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 21-25, September 1999, Florence, Italy, Vol. 3871, pp. 325-336.



Figure 11. Image SAR du système AirSAR (bande L polarisation HV) représente une scène d'étude de 250x350 pixels d'une région agricole située près de Feltwell en Royaume Uni. (tous droits réservés à *Hunting Technical Services Ltd. (UK)*).

Méthodes	Température initiale	Energie initiale	Energie finale	Temps d'exécution
Recuit simulé	5	782845.5	541049	15 min
Recherche tabou	5	782845.5	576362	5 min
Programmation génétique	-	782845.5	575940	1 heure 35 min

Tableau 1. Température initiale, Energie initiale, Energie finale (le minimum atteint par la méthode), et temps d'exécution (sur Pentium II, 255 MHZ) des trois méthodes d'optimisation.

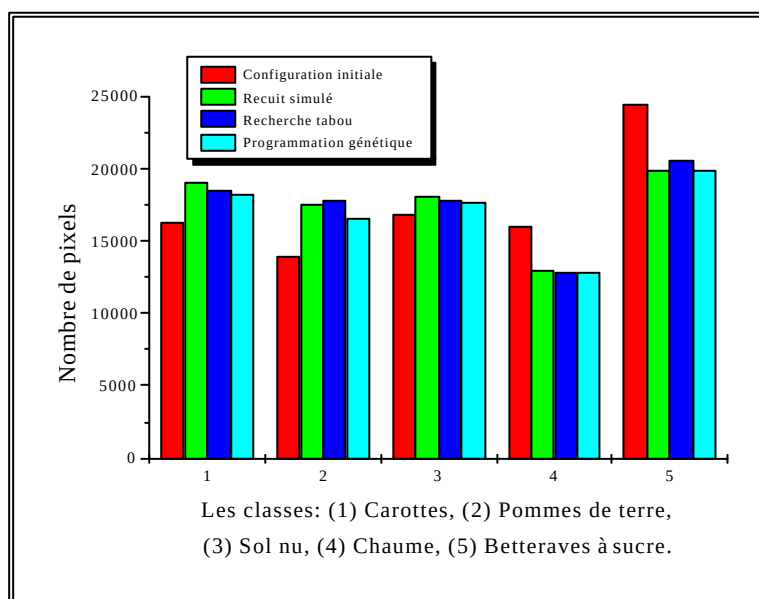


Figure 11. Nombre de pixels par classe dans les quatre images classifiées.



Figure 12. Configuration initiale.

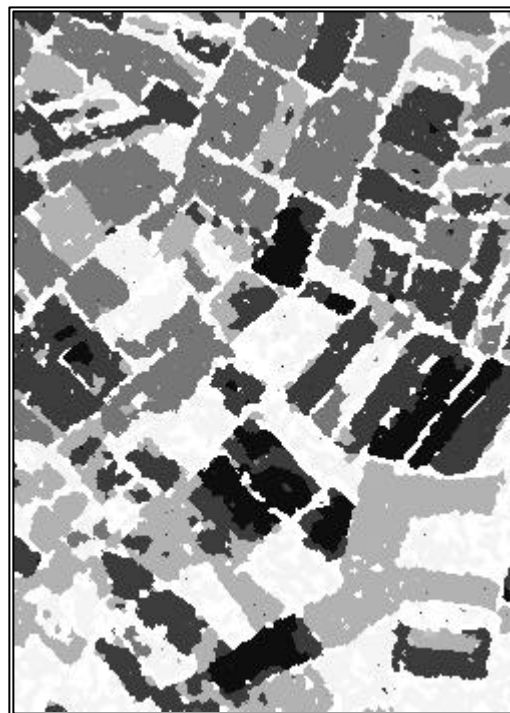


Figure 13. Recuit simulé.

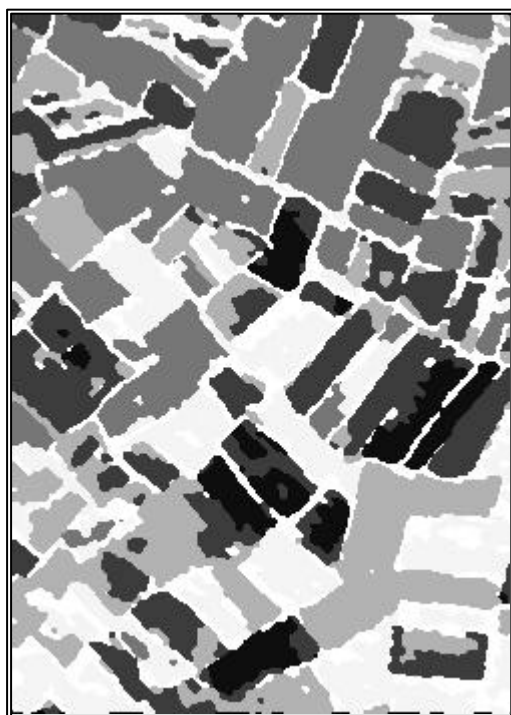


Figure 14. Recherche tabou.



Figure 15. Programmation génétique.

Chapitre 6

Fusion de Données en Télédétection

1. Introduction

Les problèmes de reconnaissance de formes tels que la classification des images de télédétection spatiale sont très complexes en général. Les images fournies par les capteurs de télédétection sont obtenues dans des conditions de prise de vue difficiles à contrôler. Elles sont bruitées en raison des effets atmosphériques, vibrations des capteurs du système d'acquisition et mécanismes d'interaction de l'onde électromagnétique. De plus, la non-disponibilité d'informations *a priori* déterministes et fiables sur l'état et les conditions de fonctionnement du système d'acquisition et sur l'état de surface de la scène imagée au moment de la prise de vue rend la classification difficile à évaluer d'une manière précise. Tout ceci rend la classification d'images de télédétection un problème de reconnaissance de forme qui est mal posé [1]. Deux approches différentes peuvent être utilisées pour résoudre ce problème. Elles se basent sur le même principe à savoir la fusion (ou combinaison) d'informations (ou de données) provenant de différentes sources et caractérisant le même phénomène physique.

En effet, les images acquises sur une même scène par différents capteurs sont partiellement redondantes puisqu'elles représentent la même scène, et partiellement complémentaires, puisque les capteurs ont des caractéristiques différentes, et que les mécanismes d'interaction de l'onde électromagnétique et de la surface imagée diffèrent. Autrement dit, l'examen de chacune des images d'une même scène, fournies par les différents capteurs, montre qu'il existe des informations significatives propres à chacun des capteurs qui peuvent être détectées facilement dans une image et difficilement dans une autre image. Il montre aussi qu'il y a une certaine relation existant entre les images des différents capteurs. Donc, l'information fournie par chaque capteur est incomplète et imprécise. Il est intéressant alors d'une part de traiter la redondance entre les différents capteurs pour réduire l'imprécision contenue dans les données de chaque capteur, et d'autre part d'utiliser leur complémentarité, en exploitant les informations significatives

portées par chacun des capteurs, pour obtenir une interprétation qui ne pourrait être obtenue à partir des données d'un seul capteur. Ceci nécessite l'exploitation simultanée de l'ensemble des images fournies par les différents capteurs. Il s'agit ici de l'approche de fusion-classification de données brutes, qui sera donnée dans le paragraphe 3.

Plusieurs classifieurs, qui utilisent les mêmes données images d'entrée, peuvent donner des réponses différentes sur l'attribution d'un pixel de l'image à la classe qui lui correspond le plus. Ceci est dû d'une part à la qualité du contenu du pixel et d'autre part à l'erreur spécifique au classifieur. Cette erreur dépend du modèle de décision du classifieur et de la base de données utilisée pour son apprentissage. Le comportement de chaque classifieur vis-à-vis des données d'entrée à classifier est déterminé en fournissant des informations différentes sur l'état de la scène. Exploiter les différentes informations que les classifieurs fournissent en utilisant les méthodes de fusion (ou combinaison) de classifieurs peut conduire à obtenir une image classifiée plus précise. Même si le classifieur est moins performant, la connaissance de son comportement apporte une certaine information utilisable à propos de la vraie classe du pixel traité [2], [3]. En fait, il n'existe aucune méthode de classification ayant pu manifester une supériorité sur les autres dans tous les problèmes et toutes les situations [2], [3]. Ainsi, si aucune méthode de classification ne peut satisfaire entièrement nos exigences, le but de la fusion est de réduire l'erreur et augmenter la fiabilité de la classification des images. Ainsi, la fusion peut éventuellement permettre d'en cumuler les avantages sans en cumuler les inconvénients, en réduisant l'erreur et en augmentant la fiabilité de la classification. Il s'agit ici de la deuxième approche, appelée fusion de classifieurs, qui sera détaillée dans le paragraphe 4.

Afin de faciliter la compréhension de ce chapitre, l'organisation de ce dernier est donnée comme suit. Pour enlever les ambiguïtés souvent rencontrées dans la littérature concernant l'utilisation du terme fusion d'informations ou combinaison d'informations, les définitions exactes de ces termes font l'objet du paragraphe 2. D'autre part, le problème de confusion entre le terme fusion de données et le terme fusion d'informations est soulevé ainsi que le sens de la fusion de classifieurs. Le paragraphe 3 est consacré à l'approche de fusion et classification de données brutes. L'approche de fusion de classifieurs est l'objet du paragraphe 4. Nous concluons ce chapitre dans le paragraphe 5.

2. Fusion d'Informations ou Combinaison d'Informations

L'idée de fusionner (ou combiner) des informations venant de plusieurs sources et portant sur un même phénomène, est une idée très simple qui vient naturellement à l'esprit. Nous faisons souvent cette fusion dans notre vie quotidienne. Par exemple, un expert humain reçoit l'information de plusieurs sources d'entrée, l'assimile et propose des conclusions.

Cependant, il faut différencier entre le terme combinaison d'informations et fusion d'informations d'une part et entre ce dernier et le terme fusion de données d'autre part.

En fait, l'idée principale de la fusion de données est de fusionner différentes informations relatives à un problème, les informations fusionnées pouvant être de natures très variées. Ces informations peuvent être des données provenant de sources différentes ou bien présentant des caractéristiques différentes, extraites des mêmes données initiales. Pour cette raison on trouve souvent dans la littérature le terme fusion d'informations au lieu de fusion de données. Il y a eu plusieurs définitions du terme fusion d'informations. En effet, le terme est souvent confondu avec le terme fusion de données. Dans un autre sens, il a été suggéré d'utiliser le terme combinaison d'informations dans un sens beaucoup plus large que la fusion d'informations. En tout cas, ces termes définissent tout processus qui implique une opération mathématique effectuée sur au moins deux sources différentes d'informations. Ces définitions sont dissociées intentionnellement et offrent différentes interprétations.

La combinaison n'est pas définie comme un terme opposé à la fusion. Elle est simplement plus générale, elle est utilisée souvent pour décrire des processus et des méthodes d'une manière générale, sans tenir compte des détails. L'intégration peut jouer un rôle semblable bien qu'elle se réfère implicitement davantage à la concaténation (i.e. augmentant le vecteur d'observations) qu'à l'extraction d'information pertinente. Un cadre dans lequel différents termes peuvent être interprétés a été donné par Benediktsson et Landgrebe [13]. Ils identifient trois niveaux de fusion : fusion au niveau des données, fusion au niveau des caractéristiques décrivant les données et fusion au niveau de la décision. Les définitions de ces trois niveaux de fusion sont détaillées dans le chapitre 1, paragraphe 3. 5 de cette thèse. Une synthèse de l'ensemble des terminologies et définitions relatives au domaine de fusion en télédétection a été aussi effectuée par Wald [14]. D'une manière générale, les techniques de fusion (ou combinaison) des informations en traitement d'images sont divisées en systèmes répartis et centralisés selon que les conclusions tirées par les différents experts, sont fusionnées, ou que les différentes sources d'informations sont utilisées ensemble par un mécanisme d'inférence simple [15]. Ici dans notre cas, il s'agit de deux approches de fusion : fusion puis classification des images brutes fournies par les satellites et fusion des réponses des classifieurs (fusion de classifieurs) en proposant un schéma de fusion des systèmes répartis.

3. Fusion-Classification d'Images Brutes

De nombreuses méthodes de fusion de données de télédétection ont été proposées dans la littérature. L'approche la plus simple consiste à concaténer les données des différents capteurs comme si elles étaient issues d'un même capteur [41]. La difficulté d'une telle

approche réside dans l'établissement d'un modèle pour les données multisource ; en effet, la fusion par concaténation des vecteurs caractéristiques n'est plus valable lorsque les différentes sources ne peuvent être décrites par un modèle commun. Des méthodes plus sophistiquées de fusion de données statistiques ont été proposées [34], [42], [43], [44]. Mais, elles nécessitent la connaissance de nombreuses informations concernant les mesures physiques des capteurs et leurs applications sont généralement totalement supervisées.

L'approche fusion-classification d'images brutes, que nous considérons ici, est une sorte de fusion de haut niveau (voir chapitre 1, paragraphe 3.5). Des approches type fusion-classification basées sur des concepts théoriques fondés ont été proposées récemment [34]-[39]. Avant de traiter ces approches et de faire une comparaison entre elles, nous allons, tout d'abord, introduire la notion d'incertitude.

3. 1. Notion d'Incertainité

L'incertitude peut représenter deux catégories de manque d'informations [40] :

- Imprécision : la difficulté de faire des distinctions exactes et précises. En d'autres termes, un manque de netteté dans les frontières.
- Ambiguïté : manque d'information sur le choix entre plusieurs alternatives.

Il y a trois types d'ambiguïté :

- Non-spécificité : il n'y a pas d'information concernant l'appartenance à une classe particulière.
- Dissonance : il y a des conflits dans les informations sur l'appartenance d'un élément à plusieurs classes.
- Confusion : il y a la possibilité de plusieurs choix pour la classe à laquelle appartient un élément.

En classification d'images, les notions d'imprécision et d'ambiguïté sont très particulièrement liées à la présence de pixels bruités et/ ou pixels mixtes. Par exemple, les pixels mixtes peuvent se rencontrer dans les deux cas suivants :

- Dans le cas où certains pixels sont intermédiaires entre deux classes (comme ceux qu'on trouve aux frontières de régions homogènes), ces pixels doivent être classés dans l'union des deux classes plutôt que dans une classe singleton. Ce cas peut être important quand la résolution spatiale du satellite est importante.
- Les pixels mixtes interviennent dans la modélisation d'une image mono-source à chaque fois que la dite source n'est pas capable de distinguer entre deux classes. Dans ce cas, seule l'information concernant l'union des classes indistinctes est accessible.

3. 2. Approches Type Fusion-Classification Fondées

Ces approches ont été conçues en utilisant la théorie Bayésienne [34], [35], la théorie des ensembles flous [36], [37], [38], et la théorie de l'évidence [39].

La théorie Bayésienne est capable de représenter la dissonance et la confusion mais elle n'est pas capable de représenter la non-spécificité. Autrement dit, la principale limite de la théorie Bayésienne est qu'elle ne peut prendre en compte l'ignorance. L'ignorance représente l'imprécision sur la mesure de probabilité d'un événement (une mesure est dite imprécise si la barre d'erreur qui existe sur sa valeur n'est pas nulle). L'autre inconvénient dans l'application de la formulation Bayésienne est qu'elle est très contraignante puisqu'elle nécessite la connaissance d'une grande quantité d'informations sur les classes : leur distribution, et leurs probabilités *a priori* [34], [35].

La structure mathématique des ensembles flous fournit une base mathématique pour traiter l'imprécision. D'un autre côté, la structure mathématique des mesures floues fournit un cadre général pour traiter l'ambiguïté. En fait, le formalisme général des mesures floues englobe les mesures utilisées dans la théorie Bayésienne. Le principal avantage de la fusion floue est que la théorie des ensembles flous fournit un grand nombre d'opérateurs de fusion, ayant des comportements et des propriétés différentes. Cette flexibilité dans le choix de l'opérateur permet à l'utilisateur d'adapter la fusion à ses données ; cependant, les opérateurs de fusion sont toujours choisis de façon supervisée. D'autant plus que, dans notre application, faire un choix judicieux d'un opérateur de fusion convenable est une tâche fastidieuse, car elle nécessite de faire plusieurs tests avec plusieurs opérateurs avant de s'affranchir d'un type d'opérateur [36]-[38].

La théorie de l'évidence fut introduite par Dempster dans les années 1960, puis développée par Shafer en 1976. Cette théorie, qui permet de représenter à la fois l'imprécision et l'incertitude, est la plus complète de deux théories précédentes. Les théories de probabilité et des ensembles flous sont des cas particuliers de la théorie d'évidence. De plus, l'approche de fusion-classification basée sur la théorie de l'évidence est plus souple et plus générale que l'approche Bayésienne et l'approche floue.

En effet, la théorie de l'évidence est capable de représenter tous les trois types d'ambiguïté. Dans la théorie de probabilité, la non-spécificité n'est pas représentable. Un manque d'informations sur plusieurs événements est en fait remplacé par des probabilités identiques pour ces événements. Par exemple, les deux cas suivants qui sont très différents sont traités exactement de la même manière :

- Nous ne connaissons rien sur un dé (non-spécificité).
- Nous avons examiné un dé avec les longs tests et nous avons vu que les probabilités des six cotés sont identiques (confusion).

En fait dans les deux cas, on attribut la probabilité de $1/6$ à chaque coté et notre ignorance dans le premier cas n'apparaît nulle part. Un premier avantage de la théorie de l'évidence est qu'elle est capable de modéliser l'ignorance et l'information manquante. En particulier, elle fournit des mesures explicites de l'imprécision et du conflit entre les informations provenant des différentes sources. Deux informations sont conflictuelles si elles conduisent à des conclusions ou à des décisions contradictoires. Un événement est incertain si sa probabilité n'est pas égale à 1 (ni 0 , auquel cas il serait certain de ne pas être). Cependant, la théorie des probabilités peut être vue comme une limite de la théorie de l'évidence dans le cas où il n'y a pas d'imprécision, et où seulement l'incertitude est considérée.

Un autre avantage de la théorie de l'évidence est la possibilité de considérer non seulement des singletons (i.e. des classes individuelles), mais aussi des unions de classes. Ce qui est particulièrement utile en présence de pixels mixtes pour ce qui est des problèmes de classification.

Des applications de la théorie de l'évidence ont été développées en imagerie médicale [45], en détection d'objet [46], [47] et en télédétection [48]. Cependant, même si les auteurs soulignent généralement les avantages de l'approche de Dempster-Shafer, leurs applications sont soit théoriques, soit sur de petits ensembles de données. Mis à part l'application au problème de la classification non-supervisée d'images multisource de télédétection [49], nous n'avons trouvé aucune application de la théorie de Dempster-Shafer sur le problème de la classification supervisée.

Nous avons développé dans notre article [39] une méthode de fusion-classification supervisée d'images de télédétection basée sur la théorie de l'évidence. La méthode proposée a été développée pour résoudre le problème d'insuffisance et/ ou de manque de fiabilité des données d'apprentissage. Elle est basée sur le concept d'intervalle de probabilités déterminé à partir des concepts de base de la théorie de l'évidence. Ainsi, les notions de base de la théorie de l'évidence et leurs relations avec le concept d'intervalle de probabilités sont données. Les différentes stratégies pour l'estimation de la fonction de croyance ainsi que la combinaison de croyances sont considérées en détail. La méthode a été appliquée sur des images radars multi-fréquences multi-polarisations fournies par le système SIR-C. Les résultats obtenus sont entièrement satisfaisants et promoteurs. En outre, la méthode n'exige aucune contrainte de modélisation des images. De ce fait, elle peut être appliquée en utilisant d'autres types de données, telles que les images multisources optiques et radars qui sont superposables et les données multi-dates. Ainsi, des applications importantes peuvent être envisagées en utilisant cette méthode, telles que l'étude diachronique de la scène observée, détection de changements et l'étude des structures géologiques. Le texte intégral de l'article est donné dans l'annexe.

4. Fusion de Classifieurs

Dans notre contexte, un classifieur est un système qui génère des données pouvant se trouver dans un des M états possibles (les classes représentant les objets terrestres). Les données générées par ce système dans un état i sont considérées comme appartenant à une classe w_i . Nous considérons tout système de traitement de données qui reçoit une forme x et donne des informations à propos de la classe correspondant à cette forme comme un classifieur au sens large du terme. La différence entre cette définition et la définition classique réside dans le type d'informations fournies par le classifieur. Dans la définition classique, un classifieur doit indiquer directement la classe correspondant à x , tandis que dans cette définition généralisée, toute information qui pourrait nous aider à trouver la vraie classe est considérée comme sortie d'un classifieur.

L'idée principale de la fusion de classifieurs est l'augmentation des performances. Cette augmentation de performance peut se traduire par plus de fiabilité dans les réponses, moins de rejet, ou bien les deux en même temps. En effet, de plus en plus, les chercheurs se rendent compte qu'il n'existe aucune méthode pouvant manifester une supériorité sur les autres dans tous les problèmes et toutes les situations [2], [3]. Toutes les méthodes ont leurs points forts et leurs points faibles. Si aucune d'entre elles ne peut satisfaire entièrement les exigences d'une application envisagée, l'utilisation simultanée de plusieurs méthodes en même temps peut éventuellement permettre d'en cumuler les avantages sans en cumuler les inconvénients. Dans notre cas d'application, l'augmentation de performances signifie aussi obtenir une image classifiée qui caractérise fidèlement l'état de surface de la scène imagée.

Différents schémas de fusion de classifieurs ont été proposés dans la littérature selon les critères de combinaison (voir chapitre 1, paragraphe 3.5). Ainsi, on trouve le critère de l'ordre d'exécution des classifieurs où on observe trois types de fusion : séquentielle, parallèle et série-parallèle. Dans la fusion séquentielle un ou plusieurs classifieurs sont exécutés et les résultats obtenus par ce ou ces classifieurs sont utilisés pour modifier l'exécution d'autres classifieurs. L'ordre d'exécution des classifieurs est important et s'il est changé, le résultat final peut être différent [16]. Dans la fusion parallèle, les résultats obtenus par les classifieurs n'affectent pas l'exécution des autres classifieurs. Par conséquent, l'ordre d'exécution des classifieurs importe peu. La fusion série-parallèle est un mélange des deux méthodes précédentes. Un autre critère de fusion est la fusion avec bouclage où le résultat de fusion peut à son tour être utilisé par les classifieurs et modifier ainsi leurs sorties. Ce schéma de fusion est très puissant puisque le système entier, i.e. les classifieurs et le fusionneur, devient un système dynamique qui peut être beaucoup plus performant qu'un système statique. En contre partie, les systèmes dynamiques, surtout ceux qui sont non-linéaires, sont beaucoup plus difficiles à analyser et à contrôler. Il y'a aussi le critère de fusion avec interaction entre les classifieurs où les classifieurs

interagissent entre eux, et modifient leurs résultats avant que la fusion soit réalisée. Enfin, nous citons le critère de fusion utilisant directement la forme x d'entrée en même temps que les réponses des classifieurs.

4. 1. Méthodes Proposées : Fusion de Classifieurs d'Imagerie

Dans le but d'améliorer la caractérisation de surface d'une scène imagée, nous avons proposé trois méthodes de fusion de classifieurs d'imagerie de télédétection [4]-[8]. En considérant la fusion des classifieurs comme une deuxième classification qui succède à celles des classifieurs. Le schéma de fusion, pour les trois méthodes, est parallèle, dans lequel il y a plusieurs classifieurs en parallèle et un fusionneur (Figure 1). Il n'y a pas d'interaction entre les classifieurs. Il n'y a pas de rebouclage du fusionneur vers les classifieurs. Les classifieurs sont fixes et ne changent pas. Ils peuvent utiliser les mêmes données, des données différentes ou des données partiellement différentes. Le fusionneur n'a pas accès à la forme x . Il traite les réponses des classifieurs et donne une réponse finale y après la fusion. L'optimisation du fusionneur est implicite dans la méthode de fusion.

4. 1. 1. Propriétés désirées pour une fusion optimale des classifieurs

La fusion optimale doit utiliser toutes les informations qui lui sont présentées. Ces informations peuvent être sur le comportement d'un seul classifieur ou sur le comportement collectif de plusieurs classifieurs. Par exemple, supposons qu'un classifieur classe souvent la classe w_i en tant que classe w_j . Cette information peut être utile dans certains cas, surtout lorsqu'il y a un litige entre la classe w_i et une autre classe que w_j .

La plupart des classifieurs qui sont utilisés pour une fusion ne sont pas indépendants. Il est par conséquent très important que la fusion soit valable pour les classifieurs dépendants. Il est vrai que le gain obtenu par la fusion de deux classifieurs dépendants sera probablement inférieur au gain obtenu s'ils étaient indépendants. Cependant la fusion doit être capable de donner le maximum de performance possible pour des classifieurs dépendants [20].

Un classifieur, même peu performant, apporte une certaine quantité d'informations utilisable, à propos de la vraie classe du pixel traité (sauf s'il est complètement aléatoire). La fusion optimale doit donc pouvoir utiliser cette information. L'intégration d'avis d'un classifieur peu performant ne doit pas diminuer la performance de la fusion. Par exemple la fusion linéaire de deux classifieurs, l'un performant et l'autre pas, donne des performances inférieures à celles du classifieur plus performant. Cela ne doit pas être le cas pour une fusion idéale.

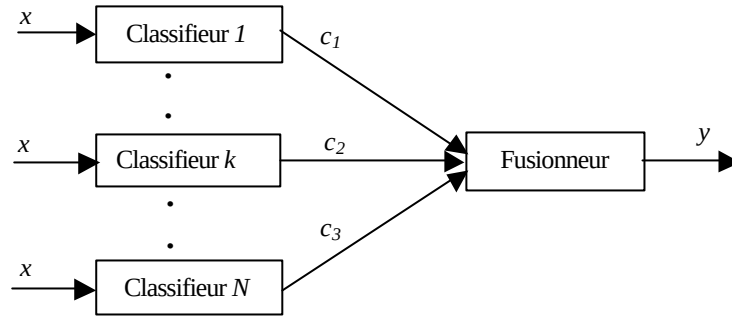


Figure 1. Cadre de travail choisi : Fusion parallèle des classifieurs d'imagerie.

Dans notre cadre de travail (figure 1), le fusionneur ne reçoit pas les données reçues par les classifieurs. Il ne voit que les sorties des classifieurs. La fusion optimale dans ce cadre doit donc être indépendante des entrées des classifieurs. Par conséquent, la méthode idéale reste la même, que les classifieurs reçoivent exactement les mêmes données ou des données différentes décrivant le même phénomène. Ceci permettrait d'utiliser la méthode optimale dans le cadre de la fusion de données. Aussi, le fusionneur ne connaît que les sorties des classifieurs et n'a pas accès à leurs entrées. Ces sorties constituent donc les seules informations à sa disposition sur la valeur de la classe à trouver, et à partir desquelles il doit élaborer sa classification.

4. 1. 2. Indépendance des classifieurs

On trouve souvent dans la littérature l'idée intuitive suivante : pour fusionner des classifieurs, il vaut mieux qu'ils ne soient pas corrélés. Il faut donc choisir des classifieurs indépendants [8], [17]. Tout d'abord, il faut clairement distinguer les notions de corrélation et dépendance et définir l'indépendance statistique des classifieurs. Dans notre cas, nous avons considéré l'indépendance conditionnelle des classifieurs qui est très importante dans les concepts des trois méthodes proposées.

4. 1. 2. 1. Dépendance et corrélation :

La dépendance de deux variables aléatoires X et Y est une propriété qui ne dépend que de leur distribution conjointe de probabilité et ne dépend pas de leurs valeurs. En revanche, la corrélation de deux variables aléatoires X et Y est une propriété qui dépend à la fois de leur distribution conjointe de probabilité et de leurs valeurs. La corrélation mesure donc un type particulier et simple de dépendance qui est la dépendance linéaire. Quand deux variables sont décorrélées, on peut dire qu'elles n'ont pas de dépendance linéaire, mais on ne peut pas dire qu'elles sont indépendantes puisqu'elles peuvent avoir des dépendances non-linéaires. En fait, X et Y sont indépendants si et seulement si : $P(X,Y) = P(X)P(Y)$, ils sont non-corrélés si et seulement si : $E(X.Y) = E(X).E(Y)$.

L'indépendance de deux variables aléatoires implique leur non-corrélation mais l'inverse n'est pas vrai. Deux variables non-corrélées ne sont pas forcément indépendantes. Il y a un cas particulier où la dépendance et la corrélation deviennent identiques. C'est le cas des distributions normales. Deux variables aléatoires avec des distributions normales seront indépendantes si elles sont non-corrélées. Autrement dit, pour les distributions normales, la seule dépendance possible est la dépendance linéaire. Deux distributions normales ne peuvent pas avoir de dépendances non-linéaires. C'est pourquoi les méthodes linéaires sont si souvent utilisées pour les distributions normales car chaque fois qu'on utilise la supposition de normalité pour un problème, les solutions deviennent des solutions linéaires et simples.

4. 1. 2. 2. Indépendance statistique des classifieurs :

En utilisant la définition de l'indépendance statistique donnée par $P(X, Y) = P(X)P(Y)$, deux classifieurs Cl_1 et Cl_2 sont indépendants si et seulement si :

$$P(c_i = w_j, c_l = w_l) = P(c_i = w_j)P(c_l = w_l), \quad \forall 1 \leq j, l \leq M \quad (1)$$

où c_1 et c_2 sont les sorties de type classe des classifieurs Cl_1 et Cl_2 , respectivement, et M est le nombre de classes. Cependant, pour avoir des classifieurs indépendants dans le sens de la définition (1), il faut que les classifieurs soient des classifieurs aléatoires (taux de reconnaissance de 0.5). Autrement dit, on suppose que les classifieurs ne rejettent pas et que leurs performances soient homogènes pour les classes, i.e. les probabilités d'une classification correcte sont identiques pour les classes. Or, pour des classifieurs non triviaux, i.e. non complètement aléatoires, ne peuvent être indépendants au sens de (1).

4. 1. 2. 3. Indépendance conditionnelle des classifieurs :

Une autre possibilité d'indépendance de deux variables aléatoires X et Y est l'indépendance conditionnelle [18], conditionnée sur une troisième variable aléatoire Z telle que : $P(X, Y|Z) = P(X|Z)P(Y|Z)$. Au lieu d'exiger que les classifieurs soient indépendants, il paraît intéressant d'utiliser cette définition et d'exiger que les classifieurs soient indépendants pour chaque classe. Cela se traduit, dans le cas de deux classifieurs de type classe, par la condition suivante :

$$P(c_i = w_j, c_l = w_l | w_i) = P(c_i = w_j | w_i)P(c_l = w_l | w_i), \quad \forall 1 \leq i, j, l \leq M \quad (2)$$

L'indépendance non-conditionnelle (1) et l'indépendance conditionnelle (2) ne sont équivalentes que pour le cas trivial où au moins l'un des classifieurs est complètement aléatoire. Ainsi, il est possible que deux classifieurs soient conditionnellement indépendants, même s'ils sont très performants. Nous proposons donc que l'indépendance conditionnelle soit utilisée comme critère d'évaluation de dépendance des classifieurs.

4. 1. 3. Hypothèses

Afin de rendre notre application plus cohérente avec les fondements théoriques des trois méthodes de fusion des classifieurs, les hypothèses suivantes sont supposées vrais :

- Nous nous détachons complètement des caractéristiques internes des classifieurs à fusionner. Tout ce qui nous intéresse dans un classifieur est la sortie qu'il fournit et non la manière dont il la calcule. Il sera donc considéré comme une boîte noire. De ce fait, les trois méthodes proposées sont générales et applicables à tout type de classifieur. Le but est de réduire l'erreur, augmenter la robustesse ainsi que la fiabilité du résultat de la classification finale de la scène imagée vis-à-vis des classes estimées.
- Nous supposons que les classifieurs sont statistiquement indépendants. L'indépendance considérée ici est l'indépendance conditionnelle par rapport aux classes [18] et non pas l'indépendance non conditionnelle [19], dans laquelle les classifieurs sont considérés comme étant des classifieurs aléatoires. En fait, le nombre de classes dans notre cas est un paramètre connu pour tous les classifieurs et les paramètres statistiques des densités de probabilité des classes sont estimés en utilisant la même base d'apprentissage pour tous les classifieurs. Ce qui différencie les classifieurs est la classe affectée à chaque pixel de l'image. Par conséquent, il est plus logique de considérer que les classifieurs sont indépendants par rapport aux classes.
- La fusion est complètement insensible aux échelles de sorties des classifieurs. La seule chose qui importe est l'estimation des vraisemblances des sorties des classifieurs, ce qui est indépendant des valeurs des sorties. Nous pouvons même transformer les sorties c_k des classifieurs avec n'importe quelles fonctions strictement monotones et cela ne changera en rien la fusion obtenue.

4. 1. 4. Méthode Bayésienne de fusion des classifieurs

Dans la méthode Bayésienne de fusion des classifieurs que nous avons proposé dans [4] et [5], nous avons supposé que les classifieurs sont statistiquement indépendants par rapport aux classes. L'implémentation de la méthode nécessite la connaissance des vraisemblances des classes par rapport aux sorties des classifieurs en termes de mesures probabilistes (distributions de probabilité). Les sous paragraphes suivants décrivent la méthodologie pour obtenir ces vraisemblances et les fusionner ultérieurement en utilisant la formule de Bayes. Le choix de la règle de décision adéquate après fusion, pour une classification précise, est aussi abordé.

4. 1. 4. 1. Méthodologie :

Supposons qu'il y ait N classifieurs Cl_k . Soit W l'ensemble fini de toutes les hypothèses (ou classes w) possibles et 2^W l'ensemble des sous-ensembles de W . Chaque classifieur k reçoit

la même forme d'entrée x (un pixel de l'image à classer) et donne en sortie une réponse c_k représentant une classe de l'ensemble de toutes les classes possibles W , i.e. : $c_k = w_j, w_j \in W$. Soit w_{jk} la réponse du classifieur c_k pour la classification du pixel x . Pour toutes les classes et tous les classifieurs, la probabilité qu'une classe w_i soit la vraie classe du pixel x sachant que le classifieur k a donné la réponse w_{jk} est :

$$P(x \in w_i | c_k = w_{jk}) \quad (3)$$

où $k = 1, 2, \dots, N$ et $i = 1, 2, \dots, M + 1$. Nous cherchons la classe qui a la probabilité *a posteriori* $P(w_i | c_1 = w_{j1}, \dots, c_N = w_{jN})$ maximale. Sous l'hypothèse d'indépendance statistique des classifieurs et en utilisant la règle de Bayes, nous obtenons :

$$P(w_i | c_1 = w_{j1}, \dots, c_N = w_{jN}) = P(w_i)^{1-N} \cdot \prod_{k=1}^N P(w_i | c_k = w_{jk}) \quad (4)$$

Appliquons la règle de Bayes pour une classification avec minimum d'erreur, le résultat de la fusion est la classe qui maximise la probabilité conditionnelle $P(w_i | c_1, c_2, \dots, c_N)$. La probabilité de w_i est conditionnée à tout ce que le fusionneur connaît du problème. Les probabilités fusionnées des classes doivent être obtenues par :

$$P_f(w_i) = P(w_i | c_1, c_2, \dots, c_N) \quad (5)$$

et la classification fusionnée de minimum d'erreur doit se faire par les fonctions discriminantes suivantes:

$$g_i(w_i) = P(w_i | c_1, c_2, \dots, c_N) \quad (6)$$

Il est clair que la probabilité fusionnée (5) est complètement insensible aux échelles de sorties des classifieurs. Dans cette fusion, la seule chose qui importe est la distribution conjointe des sorties des classifieurs, ce qui est indépendant des valeurs des sorties. Nous pouvons même transformer les sorties c_k des classifieurs avec n'importe quelles fonctions strictement monotones f_k , et cela ne changera en rien la fusion obtenue.

$$P(w_i | c_1, c_2, \dots, c_N) = P(w_i | f_1(c_1), f_2(c_2), \dots, f_N(c_N)) \quad (7)$$

Cela indique également que les normalisations n'ont aucun effet et aucune nécessité dans cette méthode. Dans le cas d'une fusion linéaire, l'emploi d'un classifieur peu performant diminue les performances obtenues par un classifieur performant.

4. 1. 4. 2. Estimation des vraisemblances des classes :

Il faut maintenant estimer les paramètres présents dans l'expression (4). Nous estimons les valeurs des probabilités à partir de la matrice de confusion M_k du classifieur k . Celle ci

étant calculée à partir des résultats de classification obtenus par le classifieur sur une base de test. Un élément de cette matrice m_{ij}^k représente le nombre de pixels dans la base qui appartenant à la classe w_i et pour lesquels le classifieur k a répondu w_j :

$$m_{ij}^k = |\{x | x \in w_i, c_k(x) = w_j\}|, \quad i = 1, \dots, M, \quad j = 1, \dots, M + 1 \quad (8)$$

Lorsque l'indice j prend la valeur $M + 1$ cela signifie qu'un ou plusieurs classifieurs ont la possibilité de donner au pixel l'étiquette de la classe de rejet w_{M+1} s'il considère que les informations mises à sa disposition ne lui permettent pas de donner un résultat fiable. Dans le cas où le résultat d'un ou plusieurs classifieurs n'aurait pas une classe de rejet, les éléments de la colonne $(j + 1)$ de la matrice de confusion correspondante seront nuls. L'estimation de la probabilité $P(x \in w_i | c_k = w_{jk})$ sera donc :

$$P(x \in w_i | c_k(x) = w_{jk}) \approx \frac{|\{x | x \in w_i, c_k(x) = w_j\}|}{|\{c_k(x) = w_j\}|} = \frac{m_{ij}^k}{\sum_{i=1}^M m_{ij}^k} \quad (9)$$

et on peut calculer les valeurs :

$$P_i = \prod_{k=1}^N P(w_i | c_k = w_{jk}), \quad i = 1, 2, \dots, M \quad (10)$$

Le problème se résume maintenant à calculer $P(w_i | c_1 = w_{j1}, \dots, c_N = w_{jN})$ à partir des P_i . Une solution proposée dans [20] consiste à simplifier (4) en ignorant $P(w_i)$ et en ajoutant une normalisation pour que la somme des probabilités devienne égale à 1 :

$$P(w_i | c_1 = w_{j1}, \dots, c_N = w_{jN}) = h \cdot P_i, \Rightarrow h = 1 / \sum_{i=1}^M P_i \quad (11)$$

Dans notre application, cette solution n'a pas donné de bons résultats. Ceci se justifie par le fait que la solution est trop simpliste et ignore complètement les probabilités *a priori* des classes. Ces probabilités ont une importance particulière dans le cas de la classification des scènes imagées, car dans une scène il y a des classes plus étendues que d'autres et par conséquent la probabilité d'existence de ces classes est plus grande que les autres. Supposer que toutes les classes sont équiprobables est une supposition qui surestime l'existence des classes et par conséquent elle ne conduit pas à une classification précise.

Une autre solution plus réaliste consiste à estimer les $P(w_i)$ en utilisant la même méthode utilisée pour estimer les $P(x \in w_i | c_k = w_{jk})$. Soit alors :

$$P(w_i) = \frac{|\{x | x \in w_i\}|}{|\{x\}|} = \frac{\sum_{j=0}^{M+1} m_{ij}}{\sum_{i=1}^M \sum_{j=0}^{M+1} m_{ij}} \quad (12)$$

Une étape de normalisation est utilisée une fois que les $P(w_i | C_i = w_{j1}, \dots, C_N = w_{jN})$ sont calculées par (4) et (10) et ceci afin de satisfaire les deux contraintes suivantes sur les probabilités :

$$\sum_{i=1}^M P(w_i | C_i = w_{j1}, \dots, C_N = w_{jN}) = \sum_{i=1}^M (P(w_i))^{1-N} \cdot P_i = 1 \quad \text{et} \quad \sum_{i=1}^M P(w_i) = 1 \quad (13)$$

4. 1. 4. 3. Règles de décision après la fusion :

Une fois les probabilités $P(w_i | C_i = w_{j1}, \dots, C_N = w_{jN})$ calculées, la règle de maximum de probabilité *a posteriori* (MAP) est utilisée pour déterminer la sortie du fusionneur y :

$$y = w_j \quad \text{si} \quad P(w_j | C_i = w_{j1}, \dots, C_N = w_{jN}) = \max_{w_i \in W} P(w_i | C_i = w_{j1}, \dots, C_N = w_{jN}) \quad (14)$$

Afin d'avoir un résultat de classification finale plus fiable en échange d'accepter des rejets, la règle de décision (14) peut être reformulée en introduisant le rejet comme il est décrit dans la chapitre 4. Ainsi, les règles de décision de Maximum Quantifié (MQ), de Maximum Notoire (MN), d'Opposition Restreinte (OR), d'Opposants Faibles (OF), et toute règle constituée de deux ou plusieurs de ces règles peuvent être utilisées (voir chapitre 4, paragraphe 3. 1. 1), afin d'obtenir une meilleure estimation de la vraie classe du pixel x et une classification plus fiable.

4. 1. 5. Méthode de la théorie de l'évidence de fusion des classifieurs

La deuxième méthode de fusion des classifieurs d'imagerie, que nous avons proposé, est basée sur le principe de la fusion parallèle des systèmes répartis en utilisant la théorie de l'évidence [6], [8].

La théorie de l'évidence [21] est complètement caractérisée par une fonction appelée fonction de masse : $m : 2^W \rightarrow [0,1]$ telle que : $m(\emptyset) = 0$ et $\sum_{A \subseteq W} m(A) = 1$. Sa valeur exprime le degré de l'évidence associée à l'ensemble A , appelé aussi élément focal, celui ci ne peut être associé à aucun sous-ensemble particulier de A . A chaque fonction de masse est associée une mesure de croyance : $Bel(A) = \sum_{B \subseteq A} m(B)$. Le fusionneur considère la réponse d'un classifieur comme une évidence sur la vraie classe du pixel x . Si la réponse est un rejet, on dira que cette évidence soutient complètement l'ensemble W , mais ne soutient aucun de ses sous-ensembles, i.e. $m_k(W) = 1$ et $m_k(A) = 0$, " $A \not\subseteq W$ ". Si la réponse n'est pas un rejet, i.e. w_{jk} , le classifieur soutient la proposition $x \in \hat{w}_{ij}$.

Soient t_{rec}^k , t_{err}^k , et t_{rej}^k les taux de reconnaissance, d'erreur, et de rejet du classifieur k , respectivement. Supposons que la réponse du classifieur k est correcte avec une

confiance t_{rec}^k , qu'elle est fausse avec une confiance t_{err}^k , et comme la réponse de ce classifieur n'indique rien sur les autres classes, tout le reste de la confiance est donné à l'ensemble W , soit alors :

$$m_k(\{w_{jk}\}) = t_{rec}^k, \quad (15)$$

$$m_k(\overline{\{w_{jk}\}}) = t_{err}^k, \quad \text{avec} \quad \overline{\{w_{jk}\}} = W - \{w_{jk}\}, \quad (16)$$

$$m_k(W) = t_{rej}^k \quad \text{et} \quad m_k(A) = 0, \quad \forall A \not\subset W. \quad (17)$$

Nous estimons les valeurs de ces taux à partir de la matrice de confusion M_k du classifieur k , comme dans le cas de la méthode Bayésienne. L'estimation des différents taux est alors donnée par :

$$t_{rec}^k = \frac{\sum_{i=1}^M m_{ii}^k}{\sum_{i=1}^M \sum_{j=0}^M m_{ij}^k}; \quad t_{err}^k = \frac{\sum_{i=1}^M \sum_{j=0, j \neq i}^M m_{ij}^k}{\sum_{i=1}^M \sum_{j=0}^M m_{ij}^k}; \quad t_{rej}^k = \frac{\sum_{i=1}^M m_{i0}^k}{\sum_{i=1}^M \sum_{j=0}^M m_{ij}^k} \quad (18)$$

Pour obtenir une attribution de masse qui contient toutes les évidences, il faut fusionner les N attributions de masse m_k associées aux classifieurs Cl_k . Pour cela, on utilise la règle de combinaison de Dempster [21] qui exige que les évidences soient indépendantes. Pour cela il faut que les classifieurs soient statistiquement indépendants conditionnellement aux classes. Le processus de fusion des évidences en utilisant la règle de Dempster s'effectue en deux étapes. Chaque étape fusionne un nombre restreint d'évidences. La fusion dans une deuxième étape est effectuée sur les évidences résultantes.

4. 1. 5. 1. Première étape de fusion :

Dans notre application, l'espace des attributions de masse m_k est restreint aux trois éléments focaux seulement, à savoir : $\{w_j\}$, $\overline{\{w_j\}}$, et W . Toute autre hypothèse composée des hypothèses $\{w_j\}$ n'a aucune signification, puisqu'on ne peut pas considérer une classe comme un ensemble de classes de caractéristiques différentes.

La première étape de fusion consiste donc à répartir les N classifieurs en N_G groupes de classifieurs selon leurs réponses. Les classifieurs de chaque groupe ont tous la réponse w_{gi} , donc leurs fonctions d'attribution de base ont toutes les mêmes focaux, mais avec des valeurs différentes de m . Soit G_i un groupe de p classifieurs. Leurs attributions de masse $m_{k1}, m_{k2}, \dots, m_{kp}$ ont toutes trois focaux $\{w_j\}$, $\overline{\{w_j\}}$, et W . La masse totale du groupe G_i est telle que :

$$m_{Gi} = m_{k1} \oplus m_{k2} \oplus \dots m_{kp} \quad (19)$$

où $\oplus$ est la règle de combinaison de Dempster. Le calcul de m_{Gi} est récursif en utilisant la règle de Dempster à deux évidences telle que :

$$m_l = m_{kl}, \quad m_r = m_{r-1} \oplus m_{kr}, \quad r = 2, 3, \dots, p, \quad m_{Gi} = m_p \quad (20)$$

Notons que pour fusionner deux évidences la règle de Dempster est telle que :

$$m(A) = (m_l \oplus m_r)(A) = K \sum_{w_1 \cap w_2 = A} m_l(w_1) m_r(w_2) \quad (21)$$

$$K = \left(\sum_{w_1 \cap w_2 \neq F} m_l(w_1) m_r(w_2) \right)^{-1} \quad (22)$$

Une évidence de la forme $m_i(W) = 1$ n'a aucun effet sur le résultat de la fusion. En outre, nous considérons que : $0 < t_{rec}^k < 1$ et $0 \leq t_{err}^k < 1$, quelque soit le classifieur. Cette condition nous assure que les évidences fournies par les classifieurs ne seront pas contradictoires, i.e. que le dénominateur de (22) n'est pas nul. Une preuve simple est donnée dans [20]. Les cas de $t_{rec}^k = 1$, $t_{err}^k = 1$, et $t_{rej}^k = 0$ sont des cas dégénérés qui peuvent être traités à part. L'espace des attributions de masse m_{Gi} est restreint aux trois éléments focaux seulement : $\{w_j\}$, $\overline{\{w_j\}}$ et W . Le groupe G_i est considéré comme un classifieur qui a répondu w_{gi} avec : $t_{rec} = m_{Gi}(\{w_{gi}\})$, $t_{err} = m_{Gi}(\overline{\{w_{gi}\}})$ et $t_{rejet} = m_{Gi}(W)$. L'étape 1 change les N_{Cl} classifieurs en N_G classifieurs qui ont tous des réponses différentes, telles que : $w_{g1} \neq w_{g2} \neq \dots \neq w_{gN_G}$.

4. 1. 5. 2. Deuxième étape de fusion :

La deuxième étape du processus de fusion consiste à fusionner les évidences des N_G groupes telles que : $m = m_{Gi} \oplus m_{GNG} \oplus \dots \oplus m_{GNG}$. Comme les N_G groupes donnent des classes différentes les unes des autres, le calcul de m devient très lourd avec la règle de Dempster. Afin de réduire le temps de calcul des fonctions de croyance, les simplifications présentées dans [20] et [22] ont été utilisées. Sachant que $N_G \leq M$ et $N_G \leq N$, nous obtenons des relations simplifiées pour calculer les fonctions de croyance en chacune des classes :

$$\begin{aligned} a &= \sum_{i=1}^{N_G} \frac{m_{Gi}(\{w_{gi}\})}{1 - m_{Gi}(\{w_{gi}\})}, \quad b = \prod_{i=1}^{N_G} (1 - m_{Gi}(\{w_{gi}\})) \\ c &= \prod_{i=1}^{N_G} m_{Gi}(\overline{\{w_{gi}\}}), \quad F = \begin{cases} (1+a)b - c & \text{si } N_G = M \\ (1+a)b & \text{si } N_G < M \end{cases} \\ Bel(\{w_{gi}\}) &= \begin{cases} F \left(\frac{m_{Gi}(\{w_{gi}\})}{1 - m_{Gi}(\{w_{gi}\})} \cdot b + \frac{m_{Gi}(W)}{m_{Gi}(\overline{\{w_{gi}\}})} \cdot c \right) & \text{si } N_G = M \text{ ou } N_G = M - 1, i = M \\ F \left(\frac{m_{Gi}(\{w_{gi}\})}{1 - m_{Gi}(\{w_{gi}\})} \cdot b \right) & \text{sinon} \end{cases} \quad (23) \end{aligned}$$

$$Bel(\overline{\{w_{gi}\}}) = \begin{cases} F\left(\left(a - \frac{m_{Gi}(\{w_{gi}\}) - m_{Gi}(\overline{\{w_{gi}\}})}{1 - m_{Gi}(\{w_{gi}\})}\right)b - c\right) & \text{si } N_G = M \\ F\left(a - \frac{m_{Gi}(\{w_{gi}\}) - m_{Gi}(\overline{\{w_{gi}\}})}{1 - m_{Gi}(\{w_{gi}\})}\right)b & \text{si } N_G < M, i \leq N_G \\ F\left(a - \frac{m_{Gi}(\{w_{gi}\})}{1 - m_{Gi}(\{w_{gi}\})}\right)b & \text{sinon} \end{cases} \quad (24)$$

4. 1. 5. 3. Règle de décision :

Les fonctions de croyance des réponses des N_G classifieurs étant calculées, une règle de décision adéquate est utilisée pour re-classifier la scène imagée. La règle tient compte des informations portées par $Bel(\{w_{gi}\})$ et $Bel(\overline{\{w_{gi}\}})$. Soit : $Bel_{nette}(w_{gi}) = Bel(\{w_{gi}\}) - Bel(\overline{\{w_{gi}\}})$ la croyance nette en w_{gi} . La règle utilisée pour la classification finale est alors :

$$y = \begin{cases} w_j & \text{si } Bel_{nette}(\{w_j\}) = \max_{w \in W} Bel_{nette}(\{w\}) \geq a, \quad 0 < a < 1 \\ w_b & \text{sinon} \end{cases} \quad (25)$$

4. 1. 6. Méthodes de vote majoritaire de fusion de classifieurs

Notre troisième méthode de fusion de classifieurs proposée est basée sur le principe de vote [7]. Cette idée est à l'origine de nombreuses méthodes de fusion de classifieurs développées dans la communauté d'intelligence artificielle. Les plus importantes d'entre elles sont : méthodes de majorité unanime, majorité sans conflit, majorité avec seuil, majorité notoire, majorité simple, et méthode de vote pondéré. L'avantage de ces méthodes réside dans le fait que les réponses des classifieurs sont traitées directement sans qu'elles soient modélisées par des modèles probabilistes ou autres, ce qui rend le processus de fusion facile et moins coûteux.

Pour commencer, nous définissons une fonction indicatrice $I_k(w_i)$ qui prend la valeur 1 si le classifieur k a répondu la classe w_i , sinon elle prend la valeur 0. Nous supposons que les classifieurs n'ont pas tous la même importance. La réponse de chaque classifieur est pondérée par un coefficient, noté α_k , qui indique la crédibilité que nous lui accordons. Les coefficients α_k sont estimés en utilisant les mesures de séparabilité statistique entre les classes telles que la distance de Jeffries-Matisuta et la divergence transformée moyenne [23]-[26]. Pour chaque classe, le nombre de classifieurs qui ont donné cette classe est indiqué par la valeur du vote total, I_{total} , tel que :

$$I_{total}(w_i) = \sum_{k=1}^N \alpha_k \cdot I_k(w_i) \quad (26)$$

Les valeurs des votes totaux des réponses des N classifieurs étant calculées, un processus de prise de décision optimale est utilisé pour classifier le pixel en considération. Ce processus est composé de quatre étapes de décision appliquées dans l'ordre suivant.

La première étape utilise une règle de décision dite règle de majorité unanime. C'est une règle sévère qui accepte le moins de risque possible. Elle consiste à affecter au pixel considéré la classe w_i si tous les classifieurs donnent la réponse w_i , sinon la réponse finale de la fusion est le rejet :

$$y = \begin{cases} w_j & \text{si } I_{total}(w_j) = N \\ w_0 & \text{sinon} \end{cases} \quad (27)$$

Le pourcentage des pixels classifiés dans cette étape caractérise le consensus des classifieurs sur les vraies classes affectées aux pixels. Par conséquent, ces pixels sont considérés comme étant correctement classifiés.

La deuxième étape du processus de décision a pour but de classier les points de rejet non classifiés dans la première étape. La règle de décision utilisée est appelée règle de majorité sans conflit. C'est une règle qui accepte un peu plus de risque que la précédente. La classe est choisie si tous les classifieurs l'ont choisie sauf éventuellement ceux qui ont donné une réponse de rejet :

$$y = \begin{cases} w_j & \text{si } I_{total}(w_j) = N - I_{total}(w_0) > 0 \\ w_0 & \text{sinon} \end{cases} \quad (28)$$

Cependant, il arrive que certains classifieurs ne donnent pas de rejet mais sont en conflit avec les autres. Dans ces conditions, les pixels sont classifiés dans la classe de rejet.

La troisième étape du processus de décision est basée sur la règle de décision dite règle de majorité avec seuil. Une classe est choisie si la majorité des classifieurs donnent cette réponse et si cette majorité dépasse un certain seuil prédéfini, noté b , entre 0 et 1. Si b est supérieur à 0.5, cela signifie qu'une majorité absolue est nécessaire :

$$y = \begin{cases} w_j & \text{si } I_{total}(w_j) \geq b \sum_{k=1}^N a_k \\ w_0 & \text{sinon} \end{cases} \quad (29)$$

Si le taux de rejet reste encore élevé, on fait appel à une règle de décision de majorité notoire dans la dernière étape du processus de décision. Dans cette étape, pour être choisie comme réponse finale, la classe majoritaire doit en plus se distinguer de la deuxième classe par une différence supérieure à un certain seuil :

$$y = \begin{cases} w_j & \text{si } I_{total}(w_j) = \max_{w_i \in W} I_{total}(w_i), I_{total}(w) = \max_{w_i \in W - \{w_j\}} I_{total}(w_i), \\ & I_{total}(w_j) - I_{total}(w_i) \geq b \sum_{k=1}^N a_k \\ w_0 & \text{sinon} \end{cases} \quad (30)$$

Notons que l'exécution de toutes les étapes du processus de décision n'est pas nécessaire si le taux de pixels affectés à la classe de rejet n'est pas assez important. En effet, le passage à l'étape suivante n'est autorisé que si on a un taux de rejet qui dépasse un seuil prédéfini par l'utilisateur.

4. 2. Résultats Expérimentaux

A travers les différents tests et expériences que nous avons effectué sur des images simulées et images réelles de télédétection, optiques et radars, les trois méthodes de fusion de classifieurs qui nous avons développé dans ce chapitre donnent de meilleurs résultats en comparaison avec les résultats des classifieurs obtenus séparément. Cependant, la méthode de la théorie de l'évidence était la meilleure des trois méthodes de fusion de classifieurs [4]-[8]. Pour cette raison, nous nous limitons dans ce paragraphe uniquement à présenter et commenter les résultats de cette méthode.

Notre première expérience, pour tester et évaluer la méthode de la théorie de l'évidence de fusion de classifieurs, a été effectuée sur une image simulée (Figure 3) de taille 251×251 pixels. Cette image est obtenue en bruitant l'image de réalité de terrain simulée (Figure 2) par un bruit additif gaussien de variance égale à 551. Elle a été classifiée par trois classifieurs différents (développés dans le chapitre 4), à savoir : le classifieur de maximum de vraisemblances (MV), classifieur de minimum de distance de Mahalanobis (MDM), et le classifieur de minimum de distance Euclidienne (MDE). Le résultat de MDE (Figure 4) est le plus précis et l'image correspondante est la moins bruitée. La même remarque peut être formulée pour le résultat de MDM (Figure 5) par rapport au résultat de MV (Figure 6). L'image de ce dernier est très dégradée et loin de la réalité terrain.

Nous avons utilisé la même base d'échantillons d'apprentissage pour estimer les paramètres du modèle de décision de chaque classifieur, sans avoir appliqué l'approche de raffinement des échantillons que nous avons développé dans le chapitre 4. En effet, notre but dans cette première expérience est de prouver la robustesse de la méthode vis-à-vis des classifieurs moins performants. Donc, nous pouvons justifier les résultats des trois classifieurs par le degré de sensibilité de chacun d'eux à la taille et la qualité de la base d'apprentissage. La figure 7 représente l'image résultat de la fusion de trois classifieurs. Nous constatons une amélioration portée par l'image de fusion, avec quelques points de rejet, en comparaison avec les images des trois classifieurs. Il y' a moins de pixels mal classifiés. Malgré que l'image la plus dégradée soit incluse dans le processus de fusion, l'image résultat de fusion est moins bruitée que le meilleur résultat des trois classifieurs. La figure 8 nous donne les taux de fiabilité des trois classifieurs et celui du résultat de leur fusion. Ce taux permet de quantifier et mesurer la fiabilité de la classification. Il est défini par :

$$fiabilité = t_{rec} / (1 - t_{rejet}) \quad (31)$$

Le taux de fiabilité de la classification après fusion est supérieur aux taux des trois classifieurs. Notons que la matrice de confusion utilisée dans cette expérience est calculée à partir de l'image de réalité de terrain (Figure 3).

La deuxième expérience a été effectuée sur une image multispectrale de télédétection fournie par le satellite SPOT. La figure 9 représente la troisième bande spectrale du satellite. L'image représente une partie de la région de Blida (situé sur la plaine fertile de la zone Mitidja, Nord d'Algérie, 140x244 pixels). C'est une région principalement agricole avec l'existence de quelques zones urbaines et sol nu. Sept classes d'objets terrestres peuvent être distinguées, à savoir : zone urbain moins dense (villages isolés entre régions agricoles dans la partie haute-gauche de l'image), végétation naturelle moins dense (à gauche et un peu en bas de l'image), sol nu (principalement au milieu de l'image et dans la partie haute à droite de l'image), champs non cultivés (en bas à gauche de l'image), zone urbaine dense (une petite partie de la ville de Blida située en bas à droite de l'image), champs cultivés (au milieu de l'image et un peu en bas de l'image) et végétation naturelle dense (en haut à droite de l'image).

Les trois bandes spectrales du satellite ont été utilisées comme entrée pour les trois classifieurs, MDE, MDM, et MV. Les résultats de ces trois classifieurs sont représentés par les figures 10, 11 et 12, respectivement. Notons que l'approche de raffinement des échantillons d'apprentissage n'a pas été appliquée. La même conclusion que dans le cas de l'image simulée peut être formulée sur les précisions des images classifiées.

Le but de cette deuxième expérience est de ressortir la robustesse de la méthode de la théorie de l'évidence de fusion de classifieurs et de pouvoir fonctionner correctement dans des situations où l'image de référence caractérisant la réalité de la scène imagée n'est pas disponible. En effet, les matrices de confusion des classifieurs utilisées dans le processus de fusion sont calculées à partir des trois images à fusionner et non pas en utilisant l'image de réalité de terrain, et ceci de la manière suivante. Nous prenons à tour de rôle chacune des images impliquées dans la fusion comme image de référence et nous construisons les matrices de confusion des autres images à partir de cette image de référence. La matrice de confusion de chaque classifieur utilisée dans la fusion est la matrice moyenne des matrices de confusion calculées en utilisant les autres classifieurs comme références.

La fusion des trois classifieurs (Figure 13) a amélioré le résultat final de la classification. Nous observons moins de points isolés dans les zones homogènes de la scène. Les frontières entre les classes sont aussi bien délimitées. Puisque dans la littérature il n'y a pas assez de critères d'évaluation de classification en absence de réalité terrain, nous avons utilisé la définition de Haralick d'une bonne segmentation [27]. Ceci consiste à calculer la moyenne des taux de fiabilité calculés dans plusieurs zones homogènes. Le taux moyen de la fusion est supérieur au taux moyen des classifieurs (Figure 14).

Ainsi, la fusion des classifieurs, en utilisant la théorie de l'évidence, a permis d'obtenir une classification plus précise, dans les deux expériences, et prouver de ce fait sa robustesse et son efficacité.

5. Conclusion

Le problème de classification d'images de télédétection est mal posé en raison de l'insuffisance d'informations *a priori* déterministes et fiables sur l'état et la nature de la formation de l'image à l'instant de prise de vue. Deux voies différentes possibles peuvent être utilisées pour résoudre ce problème. Elles se basent sur le même principe à savoir la fusion d'informations.

La première est une sorte de fusion de haut niveau. Elle consiste à faire fusionner puis classer les images brutes. Nous avons développé une méthode de fusion-classification supervisée d'images de télédétection basée sur le concept d'intervalle de probabilités. Les bornes de cet intervalle sont estimées en utilisant la théorie de l'évidence. La méthode permet de résoudre le problème d'insuffisance et/ ou de manque de fiabilité des données d'apprentissage. Elle n'exige aucune contrainte de modélisation des images et par conséquent elle peut être utilisée pour d'autres applications de télédétection. En utilisant des images SAR multi-fréquences multi-polarisations, les résultats obtenus sont satisfaisants et promoteurs.

La deuxième voie consiste à exploiter les différentes informations que les classifieurs peuvent fournir sur la scène imagée en utilisant les méthodes de fusion des classifieurs. En considérant la fusion comme une deuxième classification qui succède à celle des classifieurs, le but est d'obtenir une image classifiée plus précise caractérisant fidèlement l'état de surface de la scène.

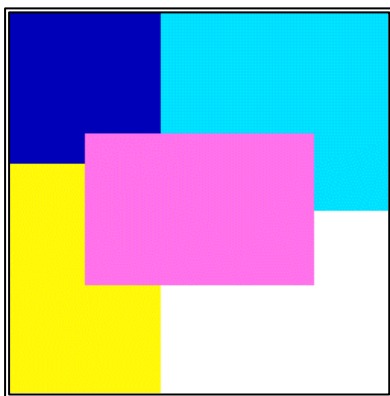
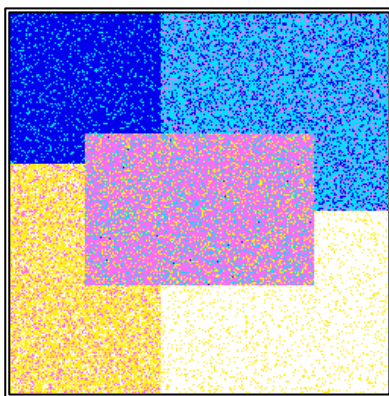
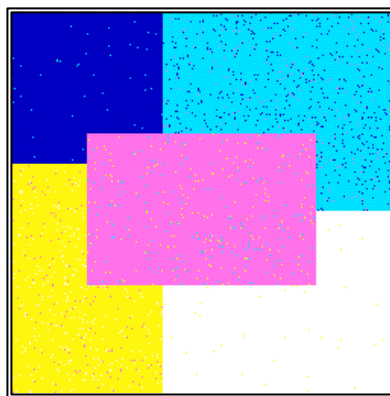
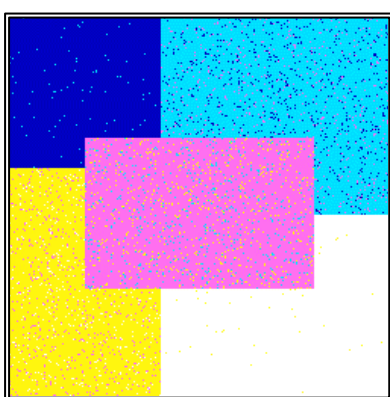
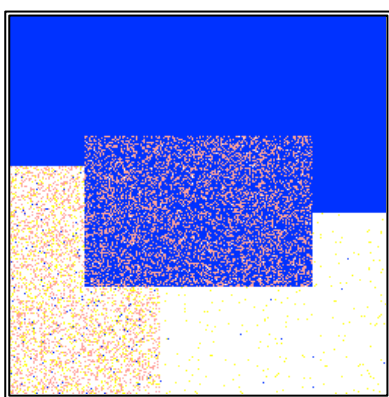
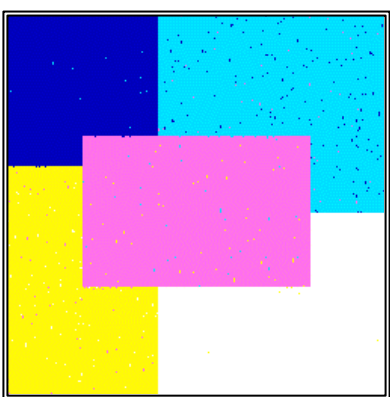
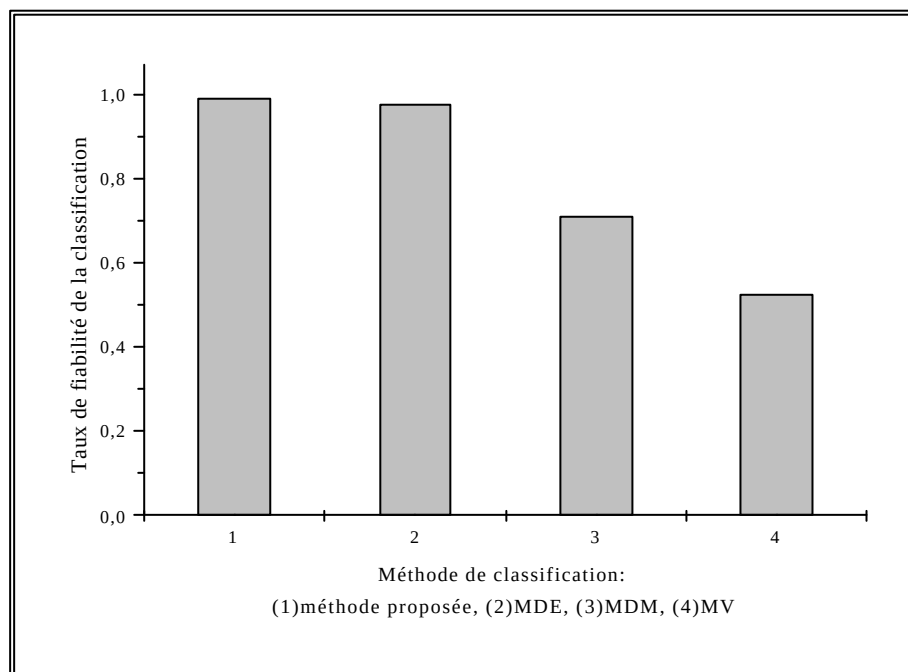
Trois méthodes de fusion des classifieurs ont été proposées. Elles cumulent les avantages de chaque classifieur sans en cumuler les inconvénients. Le schéma de la fusion des classifieurs adopté est parallèle. Il consiste à traiter les sorties des classifieurs en ignorant complètement les caractéristiques internes de ces derniers. Les trois méthodes proposées sont donc générales et applicables à tout type de classifieur. La fusion des sorties des classifieurs se base sur le principe Bayésien de fusion de données, pour la première méthode, la théorie de l'évidence, pour la deuxième méthode, et le principe de vote majoritaire, pour la troisième méthode. Ainsi, l'hypothèse d'indépendance statistique conditionnelle par rapport aux classes des classifieurs est supposée. Les trois méthodes ont été testées et évaluées en utilisant des images simulées et réelles de télédétection. Nous utilisons le taux de fiabilité de la classification comme critère d'évaluation, les résultats obtenus montrent que les méthodes de fusion de classifieurs proposées donnent de meilleurs résultats en comparaison avec ceux des classifieurs considérés séparément. De ce fait, les trois méthodes de fusion de classifieurs ont prouvé leur efficacité.

Références

- [1] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*. Winston & Sons, Washington D. C., 1977.
- [2] L. N. Kanal, "On pattern, categories, and alternate realities", *Pattern Recognition Letters*, Vol. 14, pp. 241-255, 1993.
- [3] M. Minsky, "Logical versus analogical, or symbolic versus connectionist, or neat versus scruffy", *Artificial Intelligence Magazine*, Vol. 12, No. 2, pp. 34-51, 1991.
- [4] S. Chitroub, A. Houacine, and B. Sansal, "Bayesian combination method of classifiers for cluster validation problem", In *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 21-25, September 2000, Barcelona, Spain, Vol. 3871, pp. 325-336.
- [5] S. Chitroub A. Houacine, et B. Sansal, "Fusion Bayésienne des classifieurs pour une meilleure classification des scènes imagées," In *Proceedings of International Symposium of Programming and Systems (ISPS)*, May 12-14, 2001, Algiers, Algeria, pp.69-73.
- [6] S. Chitroub, A. Houacine, et B. Sansal, "Fusion de classifieurs en utilisant la théorie de l'évidence pour l'amélioration de la classification d'images", In *Proceedings de 18^{ème} Colloque sur le Traitement du Signal et des images, (GRETSI)*, du 13 au 16 septembre 2001, Toulouse, France, pp.115-118.
- [7] S. Chitroub, A. Houacine, et B. Sansal, "Combinaison de classifieurs, une approche pour l'amélioration de la classification d'images", In *Proc. de 2^{ème} Ateliers Traitement et Analyse d'Images : Méthodes et Applications (TAIMA)*, 8-12 octobre 2001, Hammamet, Tunisie.
- [8] S. Chitroub, A. Houacine, et B. Sansal, "Fusion de classifieurs, une approche pour l'amélioration de la classification d'images", In *Proc. du 6^{ème} colloque Africain sur la Recherche en Informatique (CARI)*, du 14 au 17 octobre 2002, Yaoundé, Cameroun.
- [9] C. Nadal, R. Legault, and C. Suen, "Complementary algorithms for the recognition of totally unconstrained handwritten numerals", In *Proceedings of International Conference on Pattern Recognition*, Atlantic City, USA, pp. 443-449, 1990.
- [10] C. Suen, C. Nadal, R. Legault, T. Mai, and L. Lam, "Computer recognition of unconstrained hand-written numerals", *Proceedings of IEEE*, Vol. 80, No. 7, pp. 1162-1180, 1992.
- [11] L. Xu, A. Krzyzak, and C. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 22, No. 3, pp. 418-435, 1992.
- [12] J. Frank, "Statistical combination of multiple classifiers on image parts", In *First European Conference Dedicated to Postal Technologies (JET POSTE 93)*, Edited by SRTTP, Nantes, France, pp. 566-572, 1993.
- [13] K. Chou, L. T. Tu, and I. S. Shyu, "Performance analysis of multiple classifiers system for recognition of totally unconstrained handwritten numerals", In *Fourth International Workshop on Frontiers of Handwriting Recognition*, Taiwan, pp. 480-487, 1994.
- [14] J. A. Benediktsson and D. A. Landgrebe, "Introduction", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 37, p 1187, May 1999.
- [15] L. A. Wald, "European proposal for terms of reference in data fusion", *International Archives of Photogrammetry and Remote Sensing*, Vol. XXXII, Part 7, pp. 651-654, 1998.
- [16] L. N. Kanal, "On pattern, categories, and alternate realities", *Pattern Recognition Letters*, Vol. 14, pp. 241-255, 1993.
- [17] F. Kimura and M. Shridhar, "Handwritten numerical recognition based on multiple algorithms", *Pattern Recognition*, Vol. 24, No. 10, pp. 969-983, 1991.
- [18] B. Lamy and F. Fogelman Soulie, "Can multiple initializations improve generalization?", In *International Conference on Artificial Neural Networks*, pp. 39-44, Paris, France, 1995.
- [19] J. Pearl, D. Geiger, and T. Verma, "Conditional independence and its representations", *Kybernetika*, Vol. 25, N°. 2, 1989.
- [20] B. Moobeb, *Fusion de Classifieurs, une Nouvelle Approche*. Thèse de Doctorat, Spécialité Informatique, Ecole Polytechnique LIX, Paris Sud UFR Scientifique d'Orsay, France, 1996.
- [21] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, 1976.

-
- [22] J. A. Barnet, "Computation methods for a mathematical theory of evidence", In *17th Inter. Joint Conference on Artificial Intelligence*, pp. 868-875, Vancouver, Canada, 1985.
 - [23] S. Chitroub, *Implémentation d'un Processus d'Analyse et de Classification d'Images Satellitaires. Application à la Caractérisation des Etats de Surface*. Thèse de Magister, Numéro d'ordre 03/94-M/EL, Institut d'Électronique, Université des Sciences et Technologie Houari Boumediene, Alger, Algérie, May 1994.
 - [24] R. O. Duda, P. E. Hart, *Pattern Classification and Scene Analysis*. John Wiley & Sons, 1973.
 - [25] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. 2nd Ed., Academic Press, 1990.
 - [26] J. A. Richards, *Introduction to Remote Sensing Digital Image Analysis*. Springer-Verlag, Berlin, 1993.
 - [27] R. M. Haralick, "Statistical and structural approaches to texture", *Proceedings of IEEE*, Vol. 69, pp. 786-804, May, 1979.
 - [28] P. Bonnin, Hoeltzener-Douarin, and E. Pissaloux, "A new way of image data fusion: the multi-spectral co-operative segmentation", In *Proceedings of IEEE International Conference on Image processing (ICIP)*, Vol. 3, pp. 572-575, Washington, USA, 1995.
 - [29] C. Rosenberger and K. Chehdi, "Genetic Fusion: Application to multi-components image segmentation", In *Proceedings of IEEE International Conference on acoustic, Speech, and Signal Processing (ICASSP)*, Vol. 3, pp. 2223-2226, Istanbul, Turkey, 2000.
 - [30] J. Kittler, "Combining classifiers: A theoretical framawork", *Pattern Analysis and Application*, Vol. 1, pp. 18-27, 1998.
 - [31] J. Kittler, R. P. W. Duin, M. Hafet, and J. Matas, "On combining classifiers", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, pp. 226-239, March 1998.
 - [32] A. Petrou, A. Stassopoulou, "Advanced techniques for fusion of information in remote sensing an overview", In *Proceedings of SPIE, EUROPTO Series, Image and Signal Processing for Remote Sensing Conference*, 22-24, September 1999, Florence, Italy, Vol. 3871, pp. 264-275.
 - [33] L. O. Jiménez, M. Velez-Reyes, J. L. Rivera, and H. T. Velazquez, "Unsupervised data fusion for hyperspectral imaging", In *Proceedings of SPIE, Image and Signal Processing for Remote Sensing Conference*, 18-21 September 2001, Toulouse, France, Vol. 4541, pp. 180-189.
 - [34] A. H. Schistad Solberg, T. Taxt, A. K. Jain, "A Markov random field model for classification of multisource satellite imagery", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 34, No. 1, pp. 100-113, 1996.
 - [35] S. Chitroub, B. Sansal, "Bayesian Method for statistical classification of multisource remotely sensed images", *Second International Conference on Computational Aspects and Their Applications in Electrical Engineering (CATAEE-97)*, Philadelphia University, Faculty of Engineering, July, 22-23, 1997, Amman-Jordan, Vol.2 Part B, pp. 333-339.
 - [36] S. Chauvin, *Evaluation des théories de la décision appliquées à la fusion de capteurs en imagerie satellitaire*. Thèse de Doctorat, Université de Nantes, Nantes, France, 222 p., 1995.
 - [37] S. Chitroub, A. Houacine, and B. Sansal, "Supervised fusion-classification of multispectral images using fuzzy sets theory", *IEE International Conference on Image Processing and its Application (IPA'99)*, 13-17 July 1999, Manchester, England, pp. 96-97.
 - [38] S. Chitroub A. Houacine et B. Sansal, "Fusion et classification d'image multispectrale en utilisant la logique floue", *DAT 2000*, du 21 au 23 mai, 2000, Reghai, Algérie pp.57-63.
 - [39] S. Chitroub, A. Houacine, and B. Sansal, "Evidential reasoning-based classification method for remotely sensed images", In *Proceedings of SPIE, Image and Signal Processing for Remote Sensing Conference*, 18-21 September 2001, Toulouse, France, Vol. 4541, pp. 340-351.
 - [40] G. Klir and T. Folger, *Fuzzy Sets, Uncertainty and Information*. Prentice-Hall, 1988.
 - [41] D. G. Leckie, "Synergism of synthetic aperture radar and visible/infrared data for forest type discrimination", *Photogrammetric Eng. & Remote Sensing*, Vol. 56, pp. 1237-1246, 1990.
 - [42] A. H. Schistad Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: Fusion of Landsat-TM and SAR images", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 32, No. 4, pp. 768-778, 1994.
 - [43] I. Kanellopoulos, G. G. Wilkinson, and A. Chiuderi, "Land cover mapping using combined Landsat-TM imagery and textural features from ERS-1 synthetic aperture radar imagery", *International Society of Optical Engineering*, Vol. 2315, pp. 332-341, September 1994.
-

-
- [44] J. A. Benediktsson and P. H. Swain, "A method of statistical multisource classification with mechanism to weight the influence of the data source", In *Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS)*, Vancouver Canada, 10-14 July 1989, Vol. 2, pp. 517-520.
 - [45] I. Bloch, "Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volum effect into account", *Pattern Recognition Letter*, 1996.
 - [46] H. Rasoulia, W. E. Thompson, L. F. Kazda, and R. Parra-Loera, "Application of the mathematical theory of evidence to the image cueing and image segmentation problem", *International Society of Optical Engineering*, Vol. 1310, pp.199-206, 1990.
 - [47] J. van Cleynenbreugel, S. A. Osinga, F. Fierens, P. Suetens, A. Oosterlinck, "Road extraction from multi-temporel satellite images by an evidential reasoning appraoch", *Pattern Recognition Letters*, vol. 12, pp. 371-380, June 1991.
 - [48] T. Lee, J. A. Richards and P. H. Swain, "Probabilistic and evidential approaches for multisource data analysis", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 56, pp. 1237-1246, 1987.
 - [49] S. Le Hégarat-Masclé, I. Bloch and D. Vidal-Madjar, "Application of Dempster-Shafer evidence theory to unsupervised classification in multisource remote sensing", *IEEE Transactions on Geosciences and Remote Sensing*, Vol. 35, No. 4, pp. 1018-10301, 1997.

**Figure 2.** Image simulée.**Figure 3.** Image bruitée.**Figure 4.** Classification MDE.**Figure 5.** Classification MDM.**Figure 6.** Classification MV.**Figure 7.** Résultat de la fusion.**Figure 8.** Taux de fiabilité des trois méthodes de classification et de la fusion.

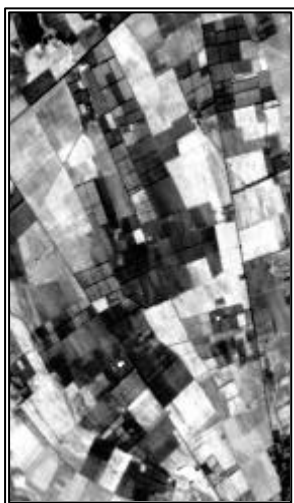


Figure 9. 3^{ème} bande spectrale du satellite SPOT (scène de Blida).

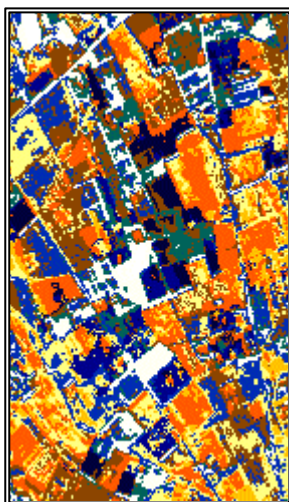


Figure 10. Classification MDE de trois bandes spectrales.

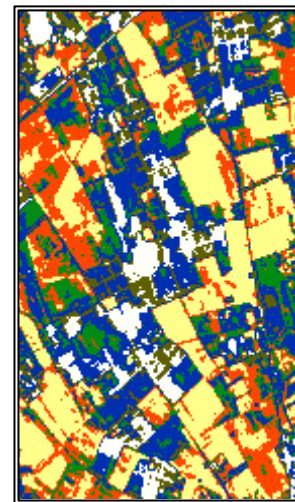


Figure 11. Classification MDM de trois bandes spectrales.

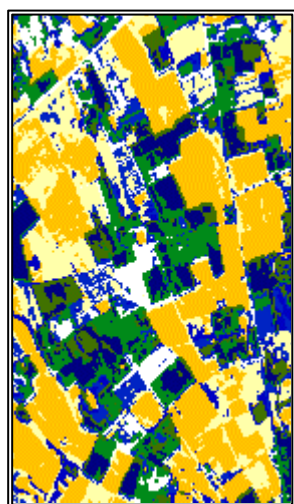


Figure 12. Classification MV de trois bandes spectrales.

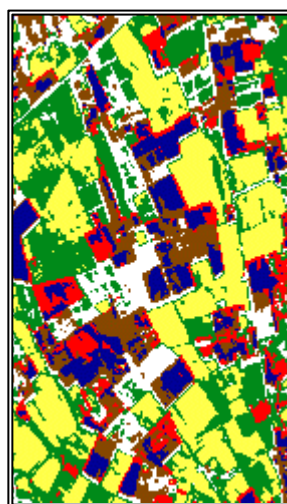


Figure 13. Résultat de la fusion de trois classifieurs.

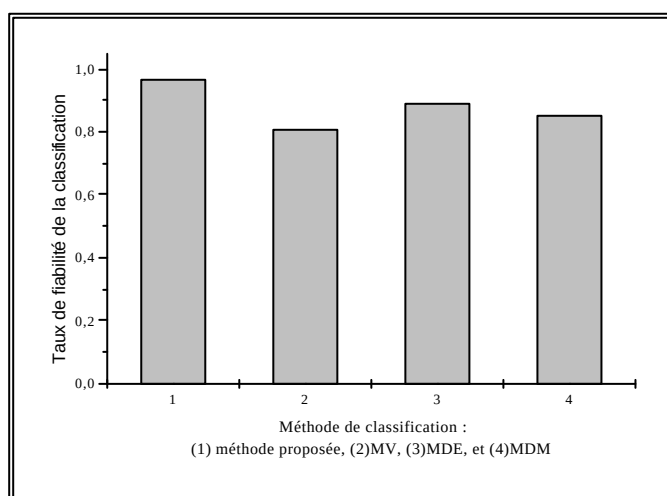


Figure 14. Taux de fiabilité des trois méthodes de classification et de la fusion.

Listes des Figures

Chapitre 1

- **Figure 1.** Spectre électromagnétique.
- **Figure 2.** Chaîne de formation de l'image discrète à partir de la scène originale.
- **Figure 3.** Image optique multispectrale fournie par le satellite.
- **Figure 4.** Chaîne simplifiée de formation de l'image optique discrète à partir de la scène originale.
- **Figure 5.** Modèle simplifié de formation de l'image optique.
- **Figure 6.** Principe de formation de l'image radar.
- **Figure 7.** Configuration du radar AirSAR.
- **Figure 8.** Illustration des résolutions en site et en azimut d'un SAR.
- **Figure 9.** Modèle simplifié de formation de l'image radar.
- **Figure 10.** Spectre de transmission de l'atmosphère terrestre.
- **Figure 11.** Erreurs géométriques introduites par les variations du relief.
- **Figure 12.** Courbes de sensibilité des capteurs SPOT.
- **Figure 13.** Analyse multidimensionnelle d'images de télédétection.
- **Figure 14.** Classification d'images de télédétection.
- **Figure 15.** Les différents types de contexte d'un pixel d'image $y(i,j)$.
- **Figure 16.** Processus de traitement et de fusion pour l'interprétation des images multisources.
- **Figure 17.** Fusion de primitives.
- **Figure 18.** Fusion séquentielle de classifieurs.
- **Figure 19.** Fusion parallèle de classifieurs.
- **Figure 20.** Fusion série-parallèle de classifieurs.
- **Figure 21.** Fusion avec bouclage de classifieurs.
- **Figure 22.** Fusion avec interaction entre les classifieurs.
- **Figure 23.** Fusion de classifieurs utilisant directement les entrées.
- **Figure 24.** Cadre de travail choisi.
- **Figure 25.** Techniques de détection de contours.

Chapitre 3

- **Figure 1.** Première bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).
- **Figure 2.** Deuxième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).

- **Figure 3.** Troisième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).
- **Figure 4.** Quatrième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).
- **Figure 5.** Cinquième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).
- **Figure 6.** Septième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tout droits réservés aux fournisseur *Eurimage*).
- **Figure 7.** Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 8.** Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 9.** Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 10.** Quatrième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 11.** Cinquième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 12.** Sixième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.
- **Figure 13.** Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 14.** Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 15.** Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 16.** Quatrième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 17.** Cinquième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 18.** Sixième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.
- **Figure 19.** Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).
- **Figure 20.** Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).
- **Figure 21.** Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).
- **Figure 22.** Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales faiblement corrélées (bandes spectrales 2, et 4).
- **Figure 23.** Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales faiblement corrélées (bandes spectrales 2, et 4).
- **Figure 24.** Septième bande spectrale du Landsat 7 bruité par un bruit additif (scène Novi Sad et Belgrade).

- **Figure 25.** Première image composante principale obtenue de la scène Novi Sad et Belgrade obtenue à partir de l'ACP à bruit ajusté des bandes spectrales bruitées.
- **Figure 26.** Première image composante principale obtenue de la scène Novi Sad et Belgrade obtenue à partir de l'ACP standard des bandes spectrales bruitées.
- **Figure 27.** Neurone simple de p entrées et une sortie.

Chapitre 4

- **Figure 1.** Les étapes d'une classification supervisée.
- **Figure 2.** Les six bandes spectrales du système de cartographe thématique aéroporté "Airborne Thematic Mapper (ATM)". (Région agricole située près de Feltwell en Royaume Uni (250x350pixel).
- **Figure 3.** Image de réalité de terrain (zones d'apprentissage sélectionnées).
- **Figure 4.** Image résultat de la classification de minimum de distance Euclidienne de six bandes spectrales d'ATM.
- **Figure 5.** Image résultat de la classification de minimum de distance de Mahalanobis de six bandes spectrales d'ATM.
- **Figure 6.** Image résultat de la classification de maximum de vraisemblance de six bandes spectrales d'ATM.
- **Figure 7.** Taux de reconnaissance dans les trois méthodes de classification.
- **Figure 8.** Taux d'erreur dans les trois méthodes de classification.

Chapitre 5

- **Figure 1.** Système de voisinage d'ordre un.
- **Figure 2.** Système de voisinage d'ordre deux.
- **Figure 3.** Organisation d'un algorithme génétique.
- **Figure 4.** Codage binaire.
- **Figure 5.** Représentation schématique du croisement uni-point.
- **Figure 6.** Représentation schématique du croisement bi-point.
- **Figure 7.** Représentation schématique du croisement uniforme (1^{ère} possibilité).
- **Figure 8.** Représentation schématique du croisement uniforme (2^{ème} possibilité).
- **Figure 9.** Représentation schématique d'une mutation dans un chromosome.
- **Figure 10.** Image SAR du système AirSAR (bande L polarisation HV) représente une scène d'étude de 250x350 pixels d'une région agricole située près de Feltwell en Royaume Uni.
- **Figure 11.** Nombre de pixels par classe dans les quatre images classifiées.
- **Figure 12.** Classification ponctuelle de l'image SAR (configuration initiale).
- **Figure 13.** Classification contextuelle de l'image SAR (recuit simulé).
- **Figure 14.** Classification contextuelle de l'image SAR (recherche tabou).
- **Figure 15.** Classification contextuelle de l'image SAR (programmation génétique).

Chapitre 6

- **Figure 1.** Cadre de travail choisi : Fusion parallèle des classifieurs d'imagerie.
- **Figure 2.** Image simulée (réalité de terrain).

- **Figure 3.** Image simulée bruitée.
- **Figure 4.** Classification MDE de l'image simulée bruitée.
- **Figure 5.** Classification MDM de l'image simulée bruitée.
- **Figure 6.** Classification MV de l'image simulée bruitée.
- **Figure 7.** Résultat de la fusion de trois classifieurs (image simulée).
- **Figure 8.** Taux de fiabilité des trois méthodes de classification et de la fusion (image simulée).
- **Figure 9.** Troisième bande spectrale du satellite SPOT (scène de Blida).
- **Figure 10.** Classification MDE de trois bandes spectrales (scène Blida).
- **Figure 11.** Classification MDM de trois bandes spectrales (scène Blida).
- **Figure 12.** Classification MV de trois bandes spectrales (scène Blida).
- **Figure 13.** Résultat de la fusion de trois classifieurs (scène de Blida).
- **Figure 14.** Taux de fiabilité des trois méthodes de classification et de la fusion (scène de Blida).

Listes des Tableaux

Chapitre 1

- **Tableau 1.** Classification des différentes dégradations subies par la scène originale dans le processus de formation d'image de télédétection optique.
- **Tableau 2.** Caractéristiques des capteurs équipant la série des satellites Landsat.
- **Tableau 3.** Caractéristiques des capteurs équipant la série des satellites SPOT.
- **Tableau 4.** Caractéristiques du système radar AirSAR.

Chapitre 3

- **Tableau 1.** Matrice de covariance des six images spectrales.
- **Tableau 2.** Matrice de transformation linéaire de l'ACP.
- **Tableau 3.** Proportion de l'information (P) en pourcentage portée par chaque image composante principale (CP) par rapport à l'information globale existante dans les images spectrales originales.
- **Tableau 4.** Rapport signal sur bruit (SNR) dans les images spectrales originales.
- **Tableau 5.** Rapport signal sur bruit (SNR) dans les images composantes principales (CP).
- **Tableau 6.** Matrice de transformation linéaire de l'ACP standardisée.
- **Tableau 7.** Proportion de l'information (P) en pourcentage portée par chaque image composante principale standardisées (CPS) par rapport à l'information globale existante dans les images spectrales originales.
- **Tableau 8.** Rapport signal sur bruit (SNR) dans les images composantes principales standardisées (CPS).
- **Tableau 9.** Statistiques des bruits additifs rajoutés aux images spectrales du satellite Landsat 7 (scène Novi Sad et Belgrade).
- **Tableau 10.** Rapport signal sur bruit (SNR) dans les images spectrales originales bruitées.
- **Tableau 11.** Rapport signal sur bruit (SNR) dans les images composantes principales (CP) de l'ACP à bruit ajusté.
- **Tableau 12.** Rapport signal sur bruit (SNR) dans les images composantes principales (CP) de l'ACP Standard.

Chapitre 4

- **Tableau 1.** Nombre d'échantillons d'apprentissage extraits des différentes parcelles sélectionnées définitivement dans les classes constituant la vérité de terrain retenue.

Chapitre 5

- **Tableau 1.** Les valeurs de la température initiale, l'énergie initiale, l'énergie finale (le minimum atteint par la méthode), et temps d'exécution (sur Pentium II, 255 MHZ) des trois méthodes d'optimisation (recuit simulé, recherche tabou et programmation génétique).

Notations

A quelques petites exceptions près, les règles suivantes sont respectées dans les notations.

Abréviations

- **ACP** : Analyse en Composantes Principales.
- **ACMD** “*Average Class Mean Difference*”: Moyenne de la difference de la moyenne de la classe.
- **ADFM** “*Average Distance From Mean*”: Moyenne des distances calculées à partir de la moyenne d la classe.
- **APEX** “*Adaptive Principal-component Extractor*” : Extracteur adaptatif des composantes principales.
- **ARSIS** : Amélioration de la Résolution Spatiale par Injection de Structures.
- **ATM** “*Airborne Thematic Mapper*” : Système de cartographe thématique aéroporté.
- **AVIRIS** “*The Airborne Visible/Infrared Imaging Spectrometer*”: Spectromètre aéroporté d’imagerie Visible/Infrarouge.
- **BLOOL** “*Bayesian Leave-One-Out likelihood*”: Probabilité de vraisemblance Bayésienne de l’échantillon enlevé sachant la moyenne échantillonnée et la covariance échantillonnée estimées à partir du reste des échantillons d’apprentissage.
- **BSA** “*Back Scattering Alignment*”: Alignement de l’onde retro-diffusée.
- **Clustering** : Terminologie anglaise signifiant groupement.
- **CNES** : Centre National d’Etude Spatial, France .
- **CP** : Composantes Principales.
- **DBFE** “*Decision Boundary-based Feature Extraction*”: Extraction de caractéristiques basée sur la limite de décision.
- **ERS** “*European Remote Sensing satellites*” : Satellites européens de télédétection.
- **H** : Polarisation Horizontale de l’onde électromagnétique émise par le radar ou réfléchié par la cible.
- **HH** : Polarisation Horizontale-Horizoantale (émission horizontale réception horizontale).
- **HIRIS** “*High Resolution Imaging Spectrometer*”.

- **HRG** “*High Resolution Geometry*”: Géométrie de haute résolution.
- **HRS** : “*High Resolution Stereoscopy*”: Stéréoscopie de haute résolution.
- **HRV** : Haute Résolution dans le Visible.
- **HRVIR** : Haute Résolution Visible et Infra Rouge.
- **HV** : Polarisation Horizontale-Verticale (émission horizontale réception verticale).
- **IHS** “*Intensity Hue Saturation*”.
- **JPL** “*Jet Propulsion Laboratory*”.
- **LOOL** “*Leave-One-Out Likelihood*”: Probabilité de vraisemblance de l'échantillon enlevé sachant la moyenne échantillonnée et la covariance échantillonnée estimées à partir du reste des échantillons d'apprentissage.
- **MAP** “*Maximum A Posteriori* ”: Classification par le maximum des probabilités *a posteriori*.
- **MDE** : Minimum de distance Euclidienne.
- **MDM** : Minimum de distance de Mahalanobis.
- **MF** “*Mean Field*”: Champ moyen.
- **MLOOL** “*Mixed-LOOLs (MLOOLs)*”: LOOL mélangés.
- **MN** : Maximum Notoire.
- **MNF** : “*Maximum Noise Fraction (MNF) transform*” : Transformation de maximum de fractions de bruit.
- **MPM** “*Maximum of the Posteriori Marginals*” : Classification de maximum des marginales *a posteriori*.
- **MODIS** “*The Moderate Resolution Imaging Spectrometer*”.
- **MQ** : Maximum Quantifié.
- **MRF** “*Markov Random Field*”: Champ aléatoire de Markov.
- **MSS** “*MultiSpectral Scanner*” : Scanner multispectral.
- **MV** : Maximum de vraisemblances.
- **NAPC** : “*Noise-Adjusted Principal Component (NAPC) transform*” : ACP à bruit ajusté.
- **NASA** “*National Aeronautics and Space Administration*”: Administration nationale d'aéronautique et de l'espace.
- **PDF** “*Probability Density Function*”: Fonction de densité de probabilité.
- **POLSAR** “*POLarimetric Synthetic Aperture Radar*”: Radar à ouverture synthétique polarimétrique.
- **OF** : Opposants Faibles.
- **OR** : Opposition Restreinte.
- **RDA** “*Regularized Discriminant Analysis*”: Analyse discriminante régularisée.
- **RLS** “*recursive least squares (RLS) computation*” : Calcul récursif au sens des moindres carrées.
- **SAR** “*Synthetic Aperture Radar*”: Radar à ouverture synthétique.
- **SIR-C/X** “*Spaceborne Imaging Radar C/X*”: Imagerie radar C/X porté sur une navette spatiale.
- **SNR** “*Signal to Noise Ratio* ” : Rapport signal sur bruit.
- **SPOT** : Système Probatoire d'Observation Terrestre.
- **TM** “*Thematic Mapper*” : Cartographe thématique.
- **V** : Polarisation Verticale de l'onde électromagnétique émise par le radar ou réfléchi par la cible.
- **VV** : Polarisation Verticale-Verticale (émission verticale réception verticale).

Symboles

Chapitre 2

- S_i : Matrice de covariance échantillonnée de la classe i .
- C : Matrice de covariance commune définie comme étant la moyenne des matrices de covariance échantillonnées des classes.
- x : Vecteur du signature spectrale d'un pixel.
- k : Nombre de bandes spectrales du satellite.
- p : Dimension d'une image multispectrale.
- w_i : Une classe i de la couverture terrestre (objet terrestre).
- m_i : Vrai vecteur moyenne de la classe i .
- $\hat{m}_i$: Vecteur moyenne de la classe i estimé.
- S_i : Vraie matrice de covariance de la classe i .
- $\hat{S}_i$: Matrice de covariance de la classe i estimée.
- M : Nombre de classes.
- $m_{i/k}$: Vecteur moyenne échantillonné de la classe i sans l'échantillon k .
- $S_{i/k}$: Matrice de covariance échantillonnée de la classe i sans l'échantillon k .
- $C_{i/k}$: Matrice de covariance commune de la classe i sans l'échantillon k .
- S_w : Moyenne des matrices de covariance échantillonnées des classes.
- L : Matrice diagonale non singulière.
- I : Matrice identité
- N : Ensemble global des échantillons d'apprentissage.
- N_i : Nombre d'échantillons de la classe i .
- α_i : Paramètre de mélange des covariances.
- l : Paramètre de mélange des covariances selon la méthode RDA.
- g : Paramètre de mélange des covariances selon la méthode RDA.
- $W\left(\frac{1}{f_i}, S_i, f_i\right)$: Distribution de Wishart centrale avec f_i degrés de liberté et S_i matrice paramètre.
- $W^{-1}((t - p - 1) Y, t)$: Distribution de Wishart inversée de paramètres Y et t pour p dimension et $t > p + 1$.
- $C_p^*(t)$: Estimateur des moindres carrés généralisé de la matrice commune.
- S : Matrice de rétro-diffusion de la cible.
- $C = S.S^{*T}$: Matrice de covariance polarimétrique de la cible.
- h : Vecteur polarimétrique complexe de la cible.
- h_i : Vecteur polarimétrique complexe du pixel i .
- q : Dimension du vecteur h .
- C_j : Matrice de covariance polarimétrique du pixel j dans une image monovue.
- $\langle C_i \rangle$: Matrice de covariance polarimétrique du pixel i dans une image multivues.
- $\langle T_i \rangle$: Matrice de cohérence du pixel i dans une image multivues.
- L : Nombre de vues (*look* dans la terminologie anglaise).
- V_m : Moyenne des matrices de cohérence des pixels appartenant à la classe w_m .
- N_m : Nombre de pixels de la classe w_m .
- $\Gamma(.)$: Fonction de Gamma.

- $P(w_m)$: Probabilité *a priori* de la classe w_m .
- $d(\langle T_i \rangle, V_m)$: Mesure probabiliste de la distance entre $\langle T_i \rangle$ et V_m .

Chapitre 3

- p : Dimension d'une image multispectrale.
- X : Vecteur des images spectrales.
- x_i : $i^{ème}$ image spectrale.
- Y : Vecteur des images composantes principales.
- y_i : $i^{ème}$ image composante principale.
- Z : Vecteur des images spectrales centrées et réduites.
- z_i : $i^{ème}$ image spectrale centrée et réduite.
- X_n : Vecteur d'images spectrales non bruitées.
- B : Vecteur du bruit additif.
- S_b^2 : Variance du bruit additif.
- S_X : Matrice de covariance des images spectrales.
- S_Y : Matrice de covariance des images composantes principales.
- S_{X_n} : Matrice de covariance des images spectrales non bruitées.
- S_B : Matrice de covariance du bruit additif.
- S_Z : Matrice de covariance des images spectrales centrées et réduites.
- C_X : Matrice de corrélation des images spectrales.
- $S^{1/2}$: Matrice diagonale dont les éléments diagonaux sont les racines carrées des écart-types des images spectrale constituant le vecteur X .
- λ_i : Valeur propre de la matrice S_X .
- d_i : Vecteurs propres normalisés associés aux valeurs propres λ_i .
- L : Matrice diagonale dont les éléments diagonaux sont les valeurs propres λ_i .
- λ_i : Valeurs propres de la matrice de covariance des images spectrales non bruitées.
- D : Matrices des vecteurs propres orthonormés
- I : Matrice identité.
- A : Matrice de la transformation linéaire de l'ACP.
- $cov(y_i, y_k)$: Covariance entre deux images composantes principales.
- R_k : Proportion de l'information portée par les images spectrales et contenue dans la $k^{ème}$ image CP.
- H : Matrice triangulaire supérieure de la décomposition de *Cholesky* de la matrice S_B .
- H^T : Matrice triangulaire inférieure de la décomposition de *Cholesky* de la matrice S_B .
- $PL(x_i)$: Prédiction linéaire de l'image spectrale x_i à partir du reste des $(p-1)$ images spectrales.
- e : Erreur au sens des moindres carrées de la prédiction linéaire.
- $r_{i(p-1)}^2$: Coefficient de corrélation multiple entre l'image spectrale x_i et le reste des $(p-1)$ images spectrales.
- U_s : Matrice de covariance du bruit additif de statistiques différentes à travers les images spectrales.

- W : Vecteur des poids synaptiques.
- b : Paramètre (ou taux) d'apprentissage.

Chapitre 4

- X : Vecteur d'images à classifier
- x : Pixel à classifier.
- $W = \{w_1, w_2, \dots, w_M\}$: Ensemble de toutes les classes possibles sans la classe de rejet.
- M : Nombre de classes.
- w_{M+1} : Classe de rejet.
- $W_+ = W \cup \{w_{M+1}\}$: Ensemble de toutes les classes possibles incluant la classe de rejet.
- $Cl : X \rightarrow W_+$: Espace des classes.
- R_{w_i} : Régions de décision.
- $C(w_i|x)$: Coût de classification de la forme x comme w_i alors que la vraie classe est w_j .
- $P(w_j|x)$: Probabilité *a posteriori* de la classe w_j .
- $P(x|w_i)$: Probabilité conditionnelle de la classe w_j .
- $P(w_j)$: Probabilité *a priori* de la classe w_j .
- $P(x)$: Densité de probabilité de x .
- m_i : Vecteur moyenne de la classe w_i .
- S_i : Matrice de covariance de la classe w_i .
- R_{Bayes} : Risque de Bayes.
- $I(i, j)$: Fonction de coût 0-1.
- $\kappa(x)$: Risque de Bayes conditionnel.
- $g_i(x)$: Fonctions discriminante de la classe w_i .
- d_{ij} : Fonction delta de Kroneker.
- T_i : Seuil correspondant à la classe w_i , appelé aussi le coût de rejet.
- $X_{w_j}^A$: Ensemble des échantillons tiré de la classe w_j .
- X^A : Ensemble des échantillons d'apprentissage constitué de tous les échantillons tirés des classes.
- c_{ij} : Eléments de la matrice de confusion.
- c_{i+} : Somme totale des éléments en lignes de la matrice de confusion.
- c_{j+} : Somme totale des éléments en colonnes de la matrice de confusion.
- $N = M \times M$: Nombre total de pixels de la matrice de confusion.
- P_A : Indicateur de précision du producteur (classifieur) de l'image classifiée (*Producer's Accuracy*).
- U_A : Indicateur de précision de l'utilisateur de l'image classifiée (*User's Accuracy*).
- O_A : Indicateur de précision globale de la classification (*Overall Accuracy*).
- $\hat{K}$: Coefficient KAPPA.

- S : Ensemble global des échantillons d'apprentissage
- S_1 : Sous-ensemble des échantillons d'apprentissage pour le classifieur.
- S_2 : Sous-ensemble des échantillons test à partir duquel sera constituée la matrice de confusion.

Chapitre 5

- $g = (S, \mathcal{X})$: Un graphe .
- $S = \{s_1, s_2, \dots, s_N\}$: Ensemble des sommets (ou des sites s) .
- $\mathcal{X}$: Ensemble des arêtes $e_{ij} \in \mathcal{X}$.
- V_s : Voisinage de s .
- $V = \{V_s | s \in S\}$: Système de voisinage pour g .
- $C \subseteq S$: Clique si chaque paire de sites dans C est voisine.
- V : Ensemble des cliques.
- L : Ensemble fini des étiquettes.
- l : Etiquette l appartenant à un ensemble fini des étiquettes L .
- w : Configuration.
- $w_s \in L$: Désigne l'étiquette au site s .
- W : Ensemble de toutes les configurations possibles.
- V : Fonction potentielle.
- $U(w)$: Fonction d'énergie sur W .
- Z : Constante de normalisation.
- $R = \{r_1, r_2, \dots, r_K\}$: Ensemble des sites.
- F : Ensemble de toutes les observations possibles.
- $X = \{X_s\}_{s \in S}$: Champ des étiquettes.
- $Y = \{Y_r\}_{r \in R}$: Champ des observations.
- $C(w, \tilde{w})$: Fonction de coût.
- $Y(a, b)$: Fonction inversible en a .
- H_r : Fonction locale définie sur un petit sous-ensemble $Y(r)$ de S .
- b : Hyperparamètre qui contrôle le degré d'homogénéité.
- J : Constante de couplage entre sites voisins.
- M : Nombre de classes.
- d : Fonction de Kroeneker.
- q : Nombre de sites dans le voisinage appartenant à la même classe du site central.
- T : Température.
- T_0 : Température initiale.
- h : Configuration perturbée.
- z : Une configuration quelconque.
- N : Nombre de pixels de l'image à classifier.
- $P_{w,h}(k-1, k)$: Probabilité de transition de l'état $k-1$ à l'état k .
- k : Caractérise un équilibre thermodynamique.
- $G_{w,h}(T)$: Probabilité de générer h à partir de w .
- $A_{w,h}(T)$: Probabilité d'acceptation de la configuration h .

- N_w : Ensemble de configurations qui ne diffèrent de la configuration w que par un élément au minimum.
- L : Longueur de la liste tabou.
- p_{si} : Probabilité de sélection.
- f_i : Qualité d'adéquation de l'individu i .
- P_c : Probabilité de croisement.
- P_m : Probabilité de mutation.
- l : Longueur de chromosome.
- pac_i : Probabilité d'adaptation cumulative de l'individus i .
- U^{wi} : Valeur de l'énergie globale de l'individus i .

Chapitre 6

- M : Nombre de classes.
- N : Nombre de classifieurs.
- $\mathbf{W}$: Ensemble fini de toutes les classes w_i possibles.
- w_i : Classe i .
- w_{M+1} : Etiquette de la classe de rejet.
- Cl_i : Classifieur i .
- c_i : Sortie du classifieur i .
- x : Un pixel de l'image.
- y : Réponse du fusionneur.
- M_k : Matrice de confusion du classifieur k .
- m_{ij}^k : Un élément de la matrice de confusion.
- $P(w_i | c_1 = w_{j1}, \dots, c_N = w_{jN})$: Probabilité *a posteriori* de la classe i .
- $P(x \in w_i | c_k = w_{jk})$: Probabilité conditionnelle du pixel x .
- $P(w_i)$: Probabilité *a priori* de la classe i .
- m : Fonction de masse.
- Bel : Fonction de croyance.
- t_{rec}^k : Taux de reconnaissance du classifieur k .
- t_{err}^k : Taux d'erreur du classifieur k .
- t_{rej}^k : Taux de rejet du classifieur k .
- N_G : Groupes de classifieurs selon leurs réponses.
- $\oplus$: Règle de combinaison de Dempster.
- $I_k(.)$: Fonction indicatrice de la réponse du classifieur k .
- α_k : Coefficient de pondération indiquant la crédibilité accordée au classifieur k .
- b : Seuil entre 0 et 1 fixant le nombre de classifieurs majoritaires.

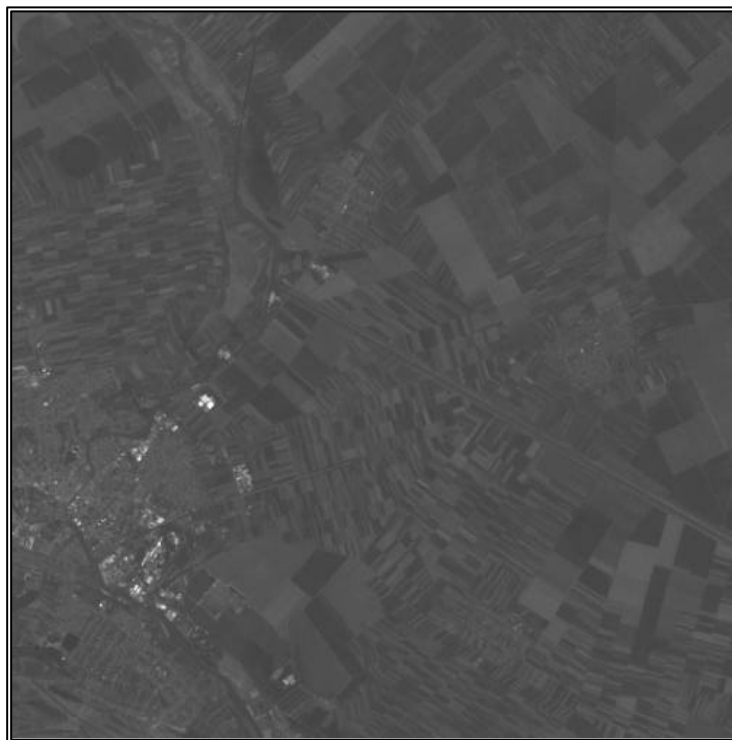


Figure 1. Première bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

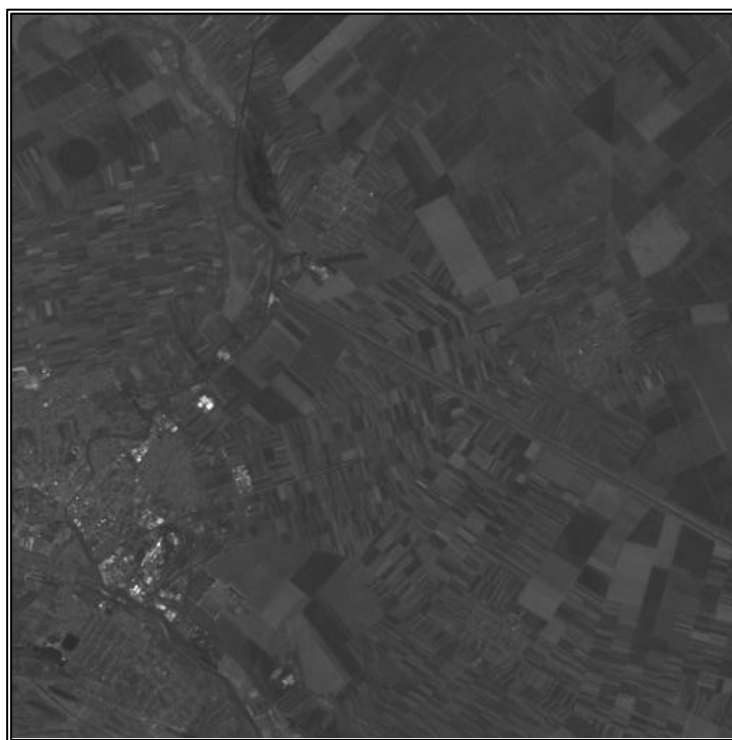


Figure 2. Deuxième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

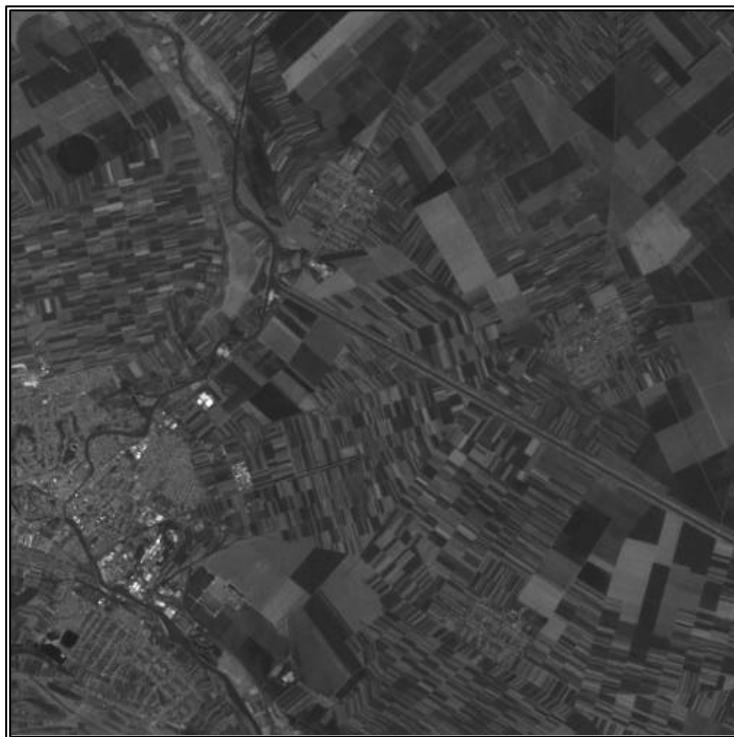


Figure 3. Troisième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

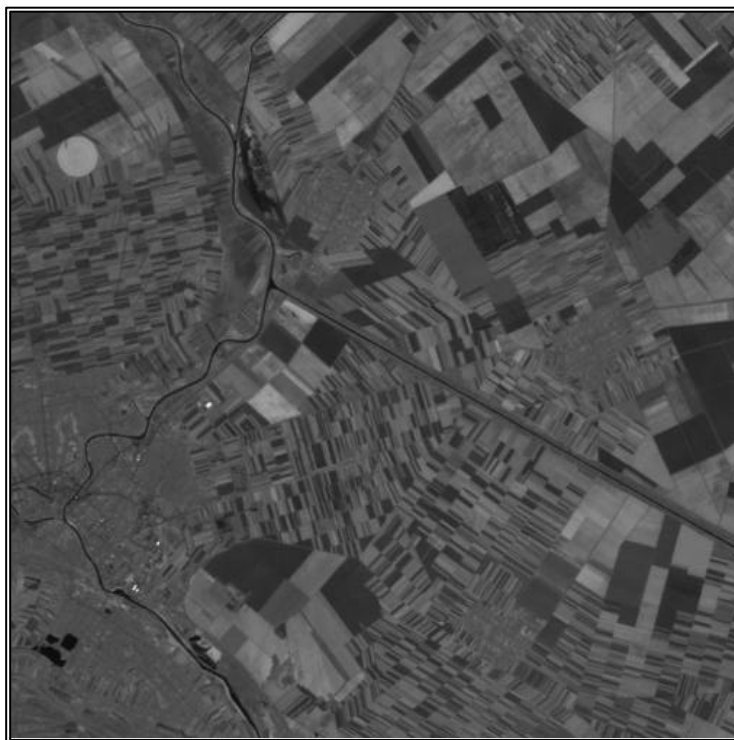


Figure 4. Quatrième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

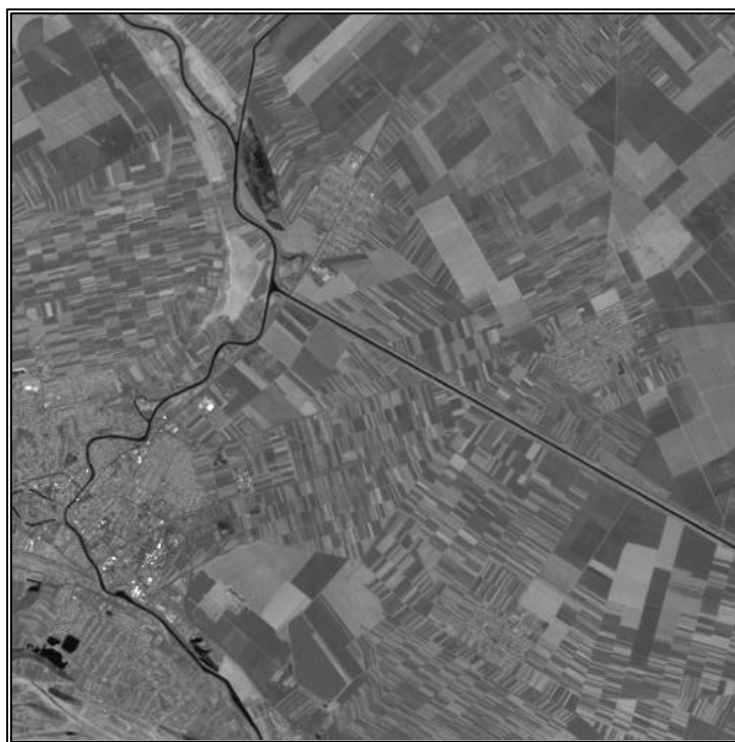


Figure 5. Cinquième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

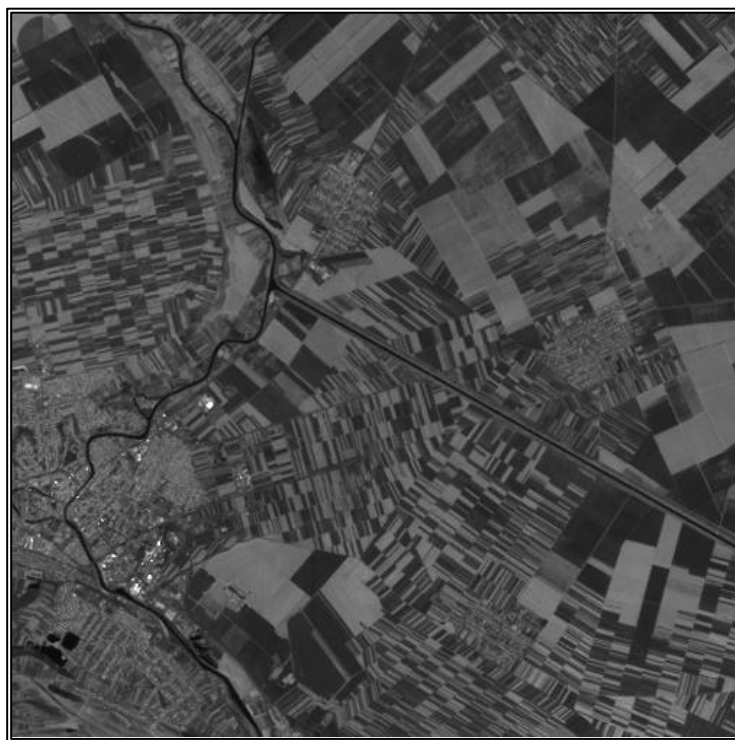


Figure 6. Septième bande spectrale du satellite Landsat-TM 7 (scène Novi Sad et Belgrade, février 2002). (Tous droits réservés au fournisseur *Eurimage*).

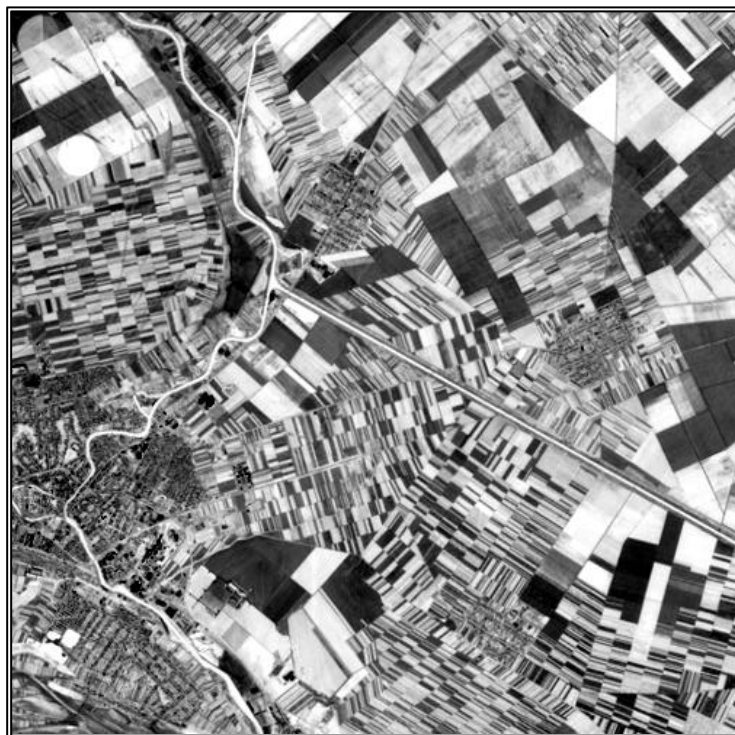


Figure 7. Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.

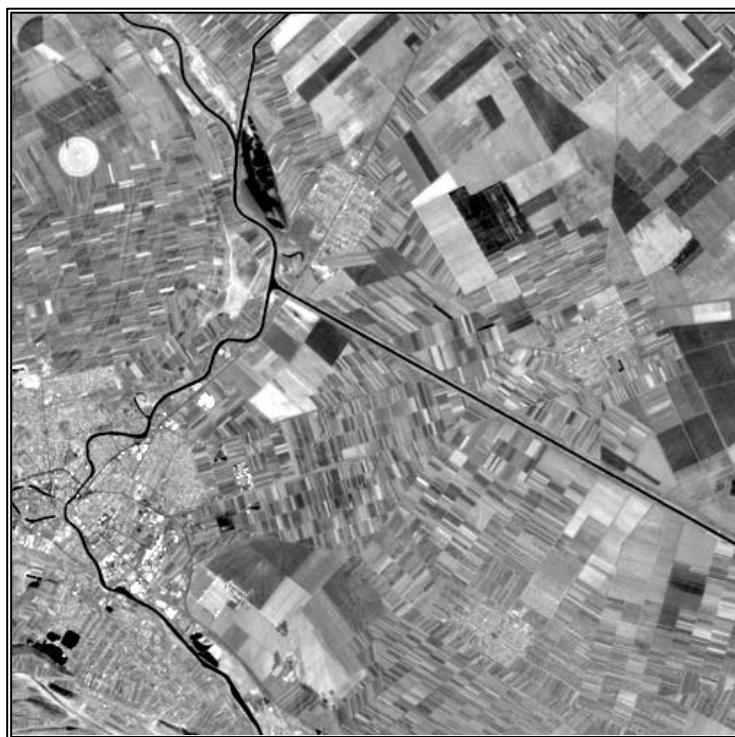


Figure 8. Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.



Figure 9. Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.



Figure 10. Quatrième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.

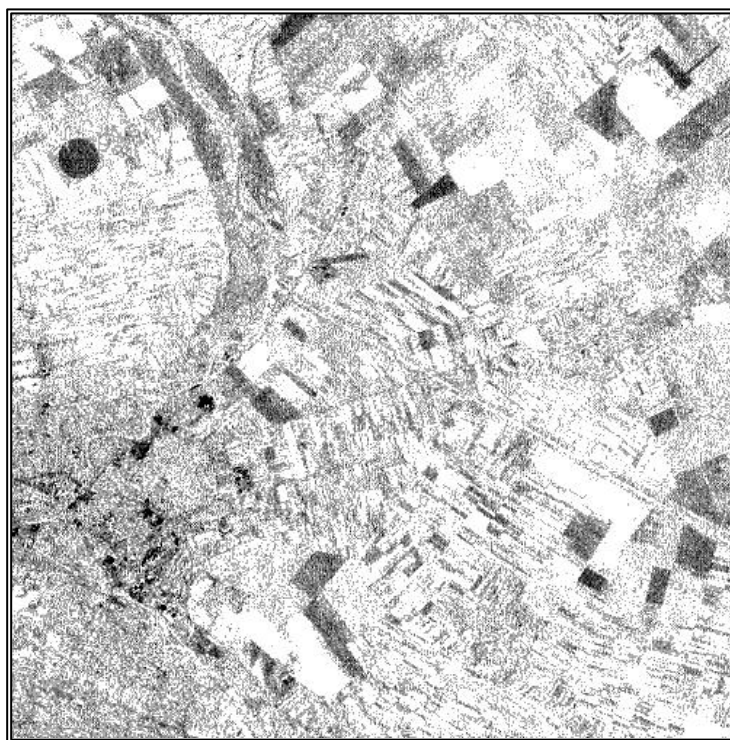


Figure 11. Cinquième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.

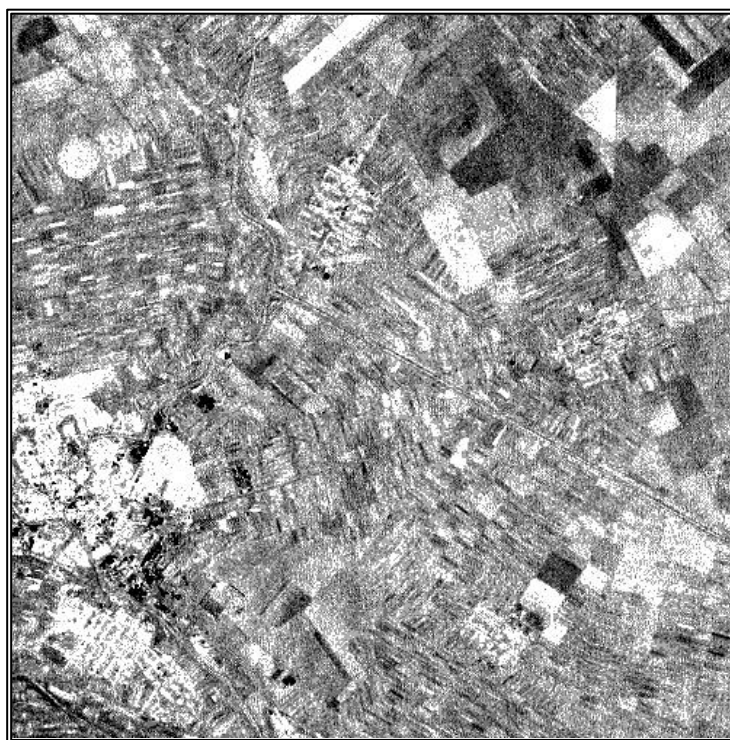


Figure 12. Sixième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP.

Tableau 1. Matrice de covariance des six images spectrales.

Bandes Spectrales	Bande 1	Bande 2	Bande 3	Bande 4	Bande 5	Bande 6
Bande 1	1.00					
Bande 2	0.921513	1.00				
Bande 3	0.934463	0.900757	1.00			
Bande 4	-0.436094	-0.226702	-0.519403	1.00		
Bande 5	0.722228	0.672519	0.839080	-0.426540	1.00	
Bande 6	0.745762	0.601454	0.835634	-0.716143	0.890394	1.00

Tableau 2. Matrice de transformation linéaire de l'ACP.

Vecteur propre 1	-0.164328	-0.155498	-0.380502	0.389510	-0.482018	-0.648055
Vecteur propre 2	0.125373	0.241934	0.238577	0.832278	0.417746	-0.040402
Vecteur propre 3	-0.369855	-0.479423	-0.568242	0.160362	0.424907	0.322802
Vecteur propre 4	0.192111	0.050595	-0.143122	0.335335	-0.607850	0.676845
Vecteur propre 5	-0.649177	0.732035	-0.179029	-0.045035	-0.028615	0.088295
Vecteur propre 6	-0.601787	-0.386014	0.650281	0.124104	-0.205670	0.090977

Tableau 3. Proportion de l'information (P) en pourcentage portée par chaque image composante principale (CP) par rapport à l'information globale existante dans les images spectrales originales.

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
P (%)	78.633169	14.814651	4.948231	1.182910	0.1616775	0.259360

Tableau 4. Rapport signal sur bruit (SNR) dans les images spectrales originales.

Bandes spectrales	Bande 1	Bande 2	Bande 3	Bande 4	Bande 5	Bande 6
SNR	61.865406	77.055611	266.741760	428.944916	423.980774	651.521790

Tableau 5. Rapport signal sur bruit (SNR) dans les images composantes principales (CP).

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
SNR	1501.980225	282.976166	94.516663	22.594753	3.088218	4.954061

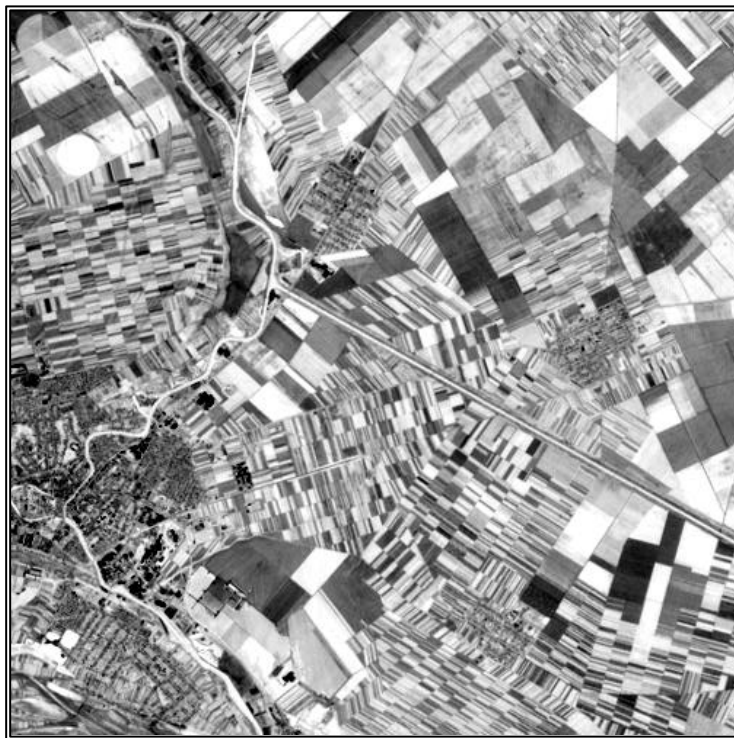


Figure 13. Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.

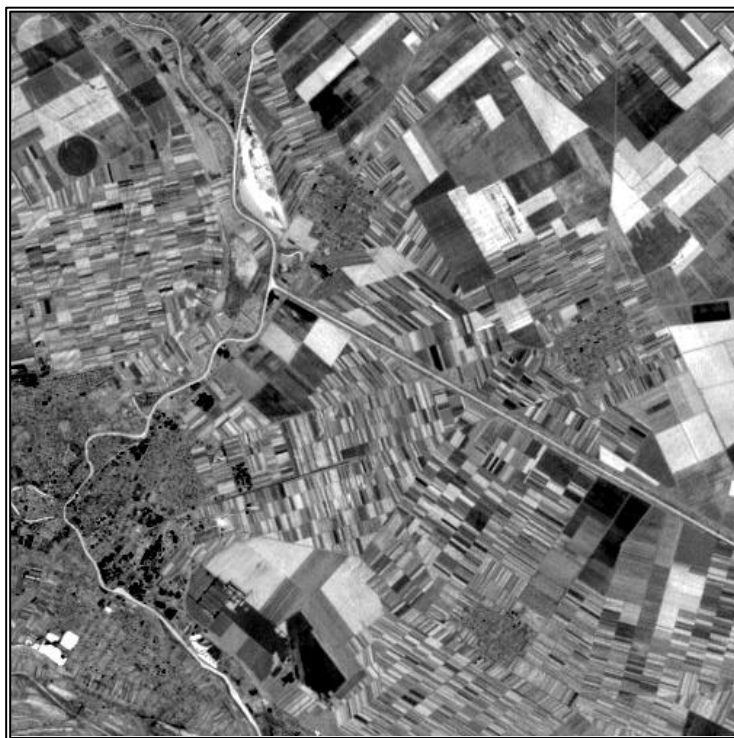


Figure 14. Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.



Figure 15. Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.

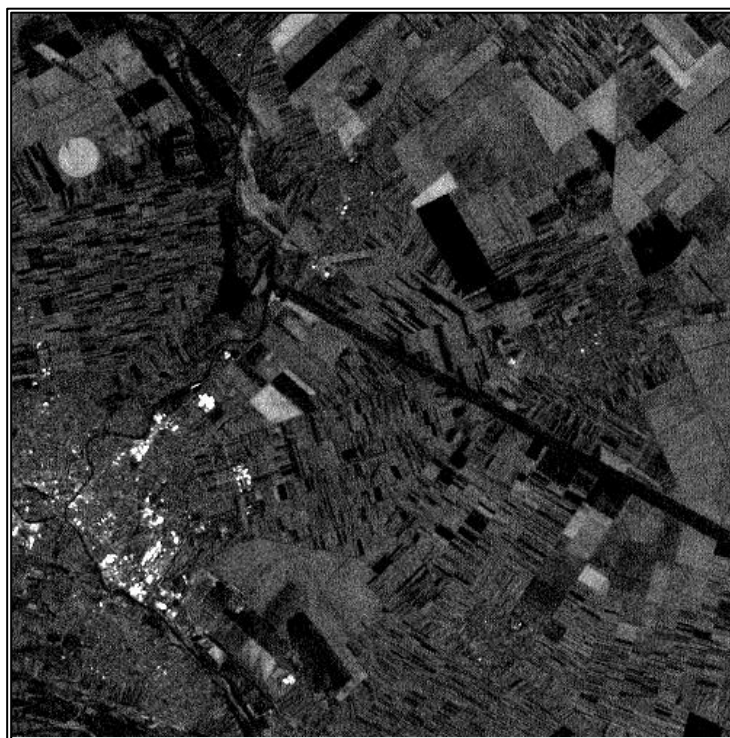


Figure 16. Quatrième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.



Figure 17. Cinquième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.

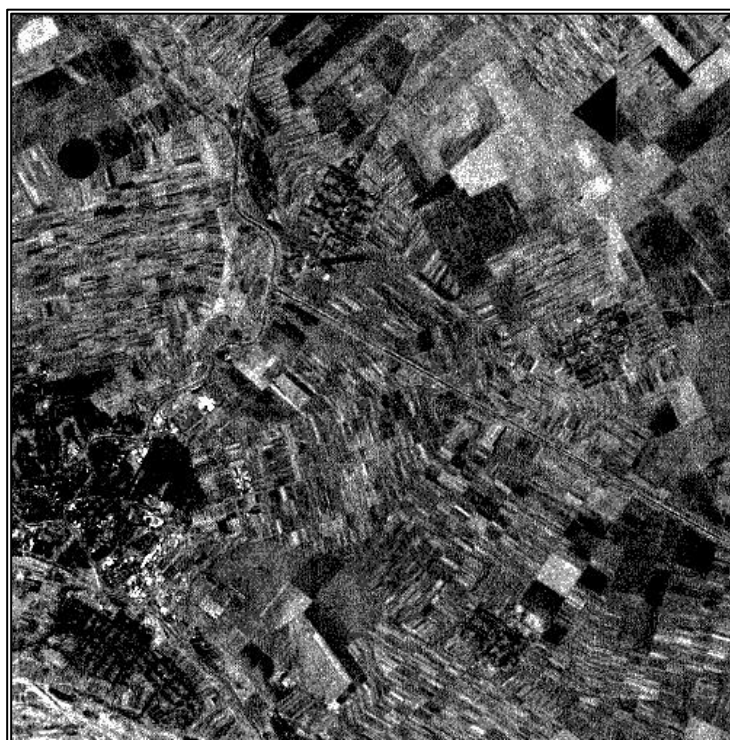


Figure 18. Sixième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de la méthode ACP standardisée.

Tableau 6. Matrice de transformation linéaire de l'ACP standardisée.

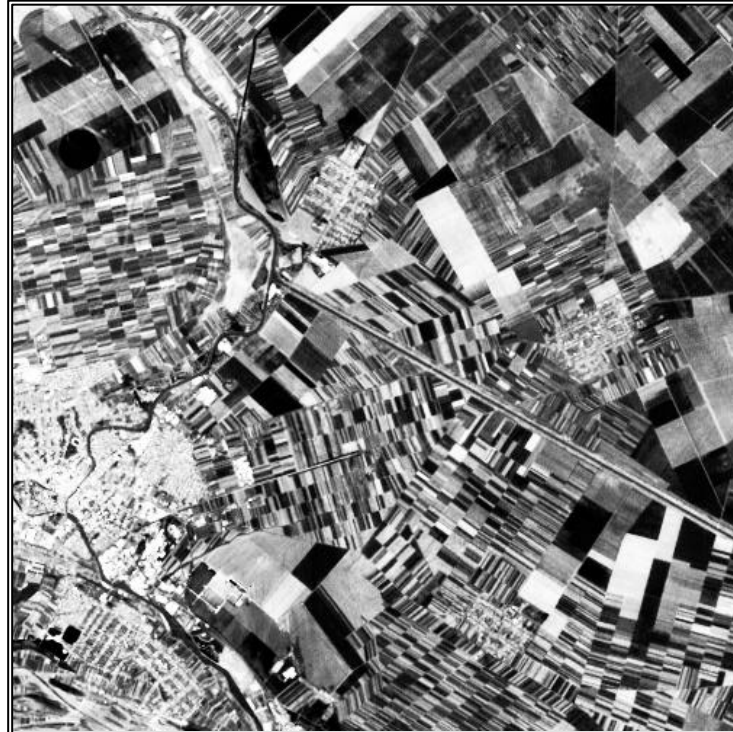
Vecteur propre 1	-0.435916	-0.400572	-0.458854	0.283866	-0.416212	-0.430301
Vecteur propre 2	-0.247122	-0.483931	-0.127383	-0.766339	0.013693	0.317887
Vecteur propre 3	-0.333675	-0.283806	-0.096248	0.444852	0.709732	0.311833
Vecteur propre 4	0.667986	-0.354250	-0.375924	0.243130	-0.217646	0.424854
Vecteur propre 5	-0.412273	0.504398	-0.236731	0.102966	-0.359515	0.616213
Vecteur propre 6	0.146271	0.380654	-0.752729	-0.254041	0.382413	-0.237338

Tableau 7. Proportion de l'information (P) en pourcentage portée par chaque image composante principale standardisées (CPS) par rapport à l'information globale existante dans les images spectrales originales.

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
P (%)	78.633169	14.814651	4.948231	1.182910	0.1616775	0.259360

Tableau 8. Rapport signal sur bruit (SNR) dans les images composantes principales standardisées (CPS).

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
SNR	4.540155	0.934564	0.401594	0.065682	0.033200	0.024804

**Figure 19.** Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).

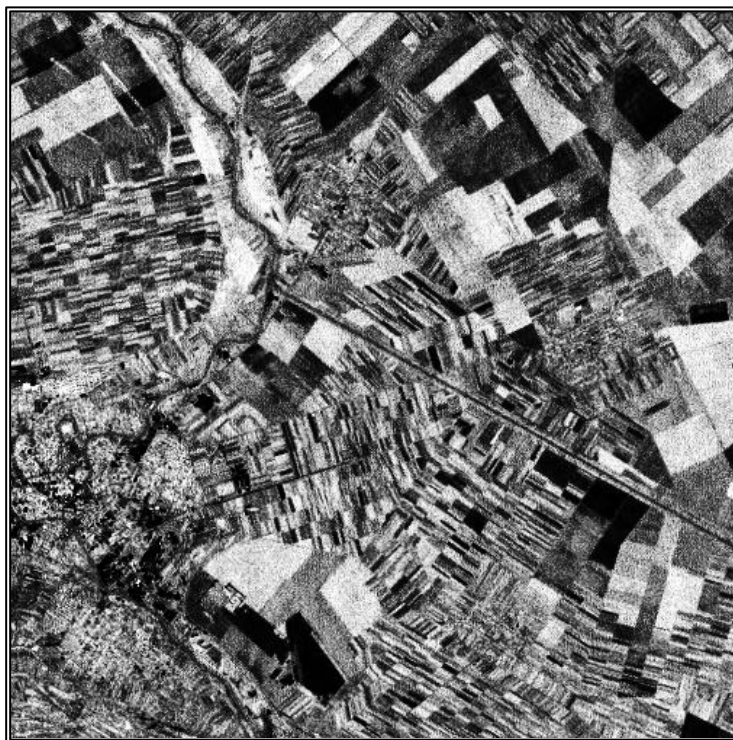


Figure 20. Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).

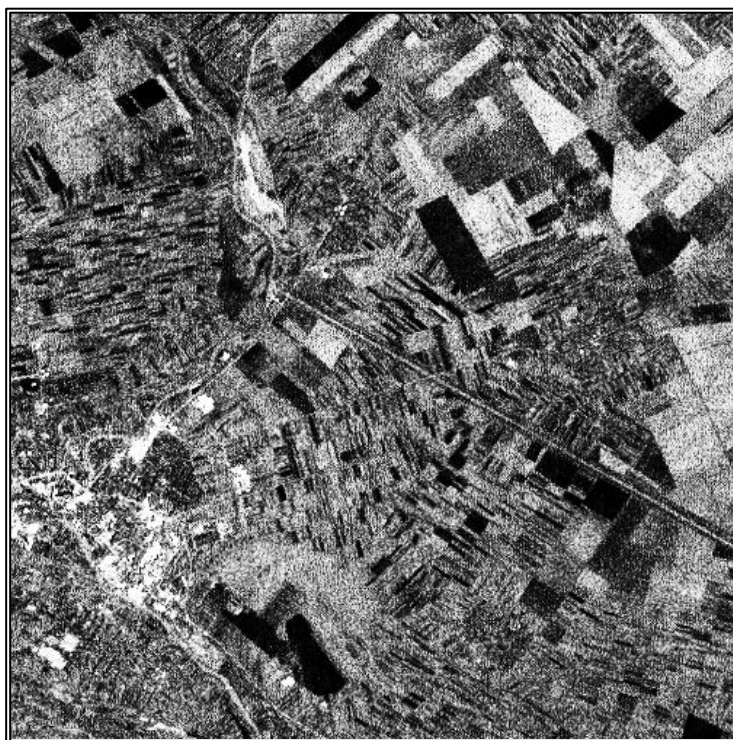


Figure 21. Troisième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales fortement corrélées (bandes spectrales 1, 2, et 3).

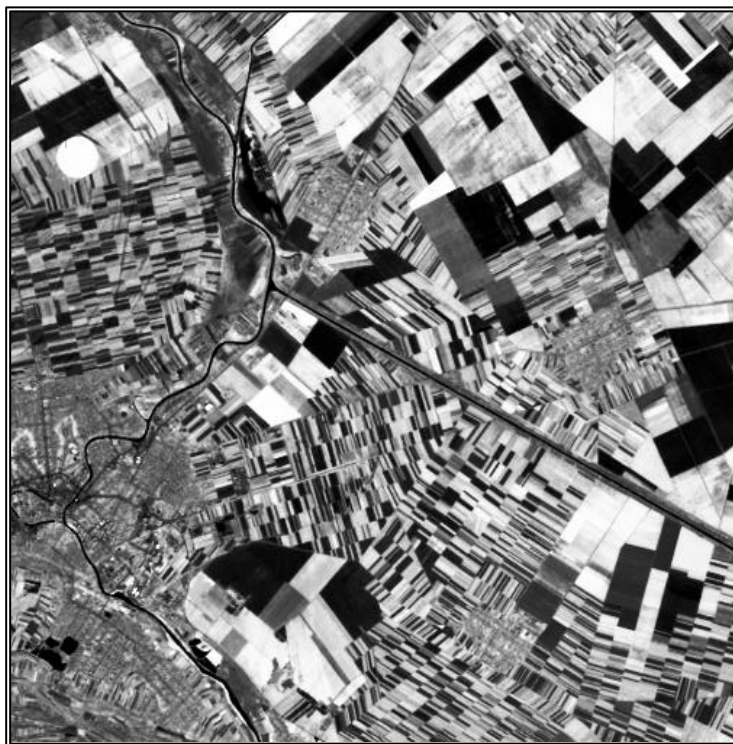


Figure 22. Première image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales faiblement corrélées (bandes spectrales 2, et 4).

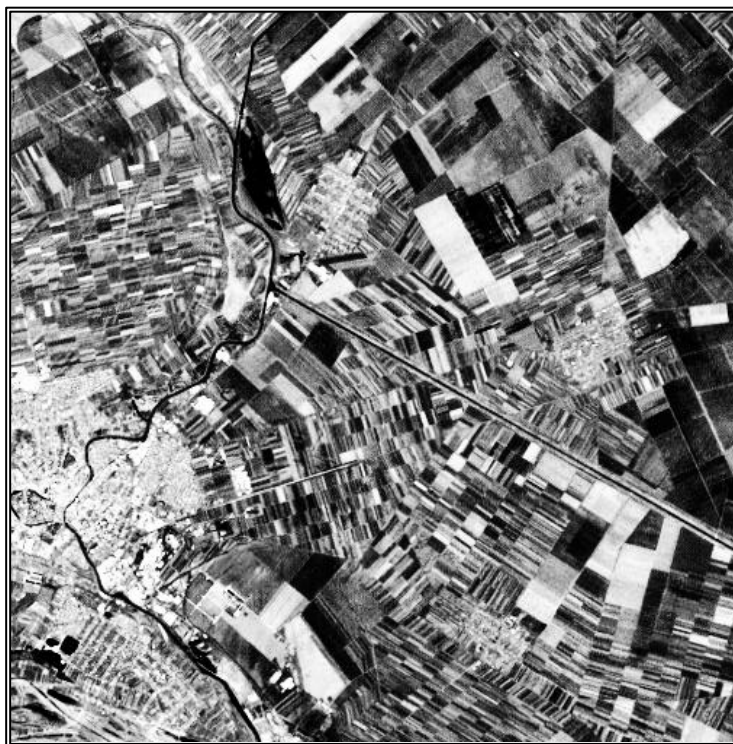


Figure 23. Deuxième image composante principale de la scène Novi Sad et Belgrade obtenue à partir de l'ACP sélective des bandes spectrales faiblement corrélées (bandes spectrales 2, et 4).

Tableau 9. Statistiques des bruits additifs rajoutés aux images spectrales du satellite Landsat 7 (scène Novi Sad et Belgrade).

Bandes Spectrales	Bande 1	Bande 2	Bande 3	Bande 4	Bande 5	Bande 6
Type de bruit et sa variance statistique	Uniforme de variance 100	Gaussien de variance 100	Uniforme de variance 151	Gaussien de variance 151	Uniforme de variance 200	Gaussien de variance 200

Tableau 10. Rapport signal sur bruit (SNR) dans les images spectrales originales bruitées.

Bandes spectrales	Bande 1	Bande 2	Bande 3	Bande 4	Bande 5	Bande 6
SNR	6.1865406	7.7055611	266.741760	428.944916	423.980774	<u>651.521790</u>

Tableau 11. Rapport signal sur bruit (SNR) dans les images composantes principales (CP) de l'ACP à bruit ajusté.

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
SNR	<u>1501.980225</u>	282.976166	94.516663	22.594753	3.088218	4.954061

Tableau 12. Rapport signal sur bruit (SNR) dans les images composantes principales (CP) de l'ACP Standard.

Images CP	CP 1	CP 2	CP 3	CP 4	CP 5	CP 6
SNR	<u>1501.980225</u>	282.976166	94.516663	22.594753	3.088218	4.954061



Figure 24. Septième bande spectrale du Landsat 7 bruité par un bruit additif (scène Novi Sad et Belgrade).



Figure 25. Première image composante principale obtenue de la scène Novi Sad et Belgrade obtenue à partir de l'ACP à bruit ajusté des bandes spectrales bruitées.



Figure 26. Première image composante principale obtenue de la scène Novi Sad et Belgrade obtenue à partir de l'ACP standard des bandes spectrales bruitées.