INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

La poursuite de projection pour la classification d'image hyperspectrale texturée

Guillaume Rellier — Xavier Descombes — Frédéric Falzon — Josiane Zerubia

N° 4152

Mars 2001

THÈME 3

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4152--FR

La poursuite de projection pour la classification d'image hyperspectrale texturée

Guillaume Rellier , Xavier Descombes , Frédéric Falzon* , Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Ariana

Rapport de recherche n° 4152 — Mars 2001 — 39 pages

Résumé : Dans ce travail, nous considérons le problème de la classification supervisée de texture à partir d'images multi-composante de télédetection, dites hyperspectrales. Ces images, le plus souvent acquises par des instruments spectro-imageurs dont le nombre de canaux est en général supérieur à 10, fournissent ainsi une représentation du paysage échantillonnée à la fois spatialement et spectralement. Le but de ce travail est de réaliser une analyse de texture qui se déroule conjointement dans ces deux espaces discrets. On recherche ainsi à enrichir la représentation "habituelle" de texture fondée sur la prise en compte des variations locales de contraste, par l'adjonction d'une connaissance sur ses variations spectrales. L'application qui est susceptible de bénéficier directement des résultats de cette étude est la classification du tissu urbain. En effet, la réponse spectrale (radiométrique) des zones urbaines est en général ambiguë du fait de la similitude de réponse spectrale de certains matériaux constitutifs du paysage urbain avec certains éléments naturels tels que l'eau, le sol nu, la végétation. La multiplication des bandes spectrales a pour conséquence de rendre plus complexes les mesures et demande également la prise en considération d'un nombre d'échantillons d'apprentissage très important. Quand le nombre de ces échantillons n'est pas suffisant, il faut passer par une étape de réduction de la dimension de l'espace d'observation. Pour prendre en compte le problème de la dimension et celui de l'analyse de texture conjointement dans le domaine spatial et spectral, on se propose ici de faire coopérer un algorithme de poursuite de projection paramétrique, déjà utilisé pour la réduction d'espace dans un cadre non-contextuel,

* Alcatel Space Industries, Cannes

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, B.P. 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : 04 92 38 77 77 - International : +33 4 92 38 77 77 — Fax : 04 92 38 77 65 - International : +33 4 92 38 77 65

à un modèle de texture par champ markovien, dit modèle markovien gaussien.

Mots-clés : image hyperspectrale, classification, texture, réduction d'espace, poursuite de projection, champs de Markov.

Projection pursuit for textured image classification

Abstract: In this work we develop a supervised texture classification algorithm for application to the class of multi-component images called hyperspectral. These images, usually recorded by spectrometers with a number of bands greater than 10, give both a spatially and spectrally sampled representation of a remote scene. The aim of this work is to perform a joint texture analysis in both discrete spaces. The use of spectral variations in this joint texture analysis scheme enables us to improve on the standard representation of textures which only takes into account the local contrast variations. A likely application of this work is urban area classification. Indeed, the spectral response of urban areas is in general ambiguous because some of its constitutive elements have the same reflectance as natural elements such as water, vegetation or bare soil. The greater number of spectral bands makes the measures more complex and so creates the need for a greater number of training samples. When the number of training samples is not sufficient, a necessary step in the analysis is to reduce the dimension of the observation space. To take into account both the problem of dimensionality and the jointly spectral and spatial texture analysis, we propose to use in cooperation a projection pursuit algorithm and a Gauss-Markov random field texture model.

Key-words: hyperspectral image processing, classification, texture, dimension reduction, projection pursuit, Markov random fields.

Table des matières

1 Introduction **6**

2 Nécessité et justification de la réduction de l'espace **8**
2.1 Nécessité 8
2.1.1 Problèmes d'estimation 8
2.1.2 Efficacité algorithmique 8
2.2 Justification 8
2.2.1 Renforcement de l'hypothèse de normalité 9
2.2.2 Corrélation 9

3 Méthodes de réduction d'espace **11**
3.1 Approches générales 11
3.1.1 Analyse en composantes principales 11
3.1.2 *Minimum Noise Fraction* 12
3.2 Approches pour la classification 13
3.2.1 Analyse discriminante 14
3.2.2 Poursuite de projection 15

4 Modélisation **20**
4.1 Modèles markoviens 20
4.2 Modèle markovien gaussien 21
4.3 Le MGMRF 21
4.3.1 Loi conditionnelle 21
4.3.2 Loi globale 23
4.3.3 Caractéristiques des espaces de grande dimension 24
4.3.4 Domaine de validité des paramètres 25
4.3.5 Estimation des paramètres 25

5 Application de la modélisation proposée à la réduction d'espace pour la classification **27**
5.1 Mise en relation de la règle de classification et de l'indice de projection 27
5.2 Problème à plusieurs classes 28
5.3 Calcul de l'indice de projection 29

6 Résultats **30**
6.1 Collage de textures urbaines 30
6.2 Résultats sur une scène AVIRIS 31
6.3 Temps de calcul 33

7 Conclusion **36**

Références **39**

1 Introduction

Une image de télédétection est une image échantillonnée spatialement, prise par un capteur survolant une scène. Les images monospectrales, ou de niveaux de gris, fournissent, pour chaque échantillon spatial (pixel), une valeur scalaire correspondant au rayonnement électromagnétique incident sur le capteur intégré sur la bande de longueur d'onde mesurée. Les images hyperspectrales, par opposition aux images monospectrales, sont des images de télédetection ayant la particularité de fournir à la fois un échantillonnage spatial et un échantillonnage spectral. Pour chaque échantillon spatial, elles ne fournissent plus une valeur scalaire, mais un ensemble de valeurs (qui peut être vu comme un vecteur), correspondant aux mesures du rayonnement faites dans un certain nombre de longueurs d'onde. Ces images constituent par conséquent un "pavé" hyperspectral dans lequel chaque point peut être indexé par ses deux coordonnées spatiales et sa coordonnée spectrale (sa longueur d'onde correspondante).

Ces images ont un nombre de canaux en général supérieur à 10, et chacun de ces canaux est associé à une longueur d'onde du spectre allant du visible à l'infrarouge. Par rapport aux images multispectrales (dont le nombre de canaux est supérieur à 1 et inférieur à 10), ce découpage spectral plus fin autorise des mesures radiométriques plus précises, et permet donc une caractérisation plus exacte du matériau présent sur le terrain. Ceci autorise en particulier des applications liées à la reconnaissance de sol en géologie, à l'aide à l'agriculture de précision, etc...

En contrepartie, ces mesures très précises spectralement sont peu précises spatialement. En effet, puisque les bandes spectrales sont très étroites, il faut intégrer le rayonnement sur des cellules plus grandes pour obtenir l'énérgie suffisante à la mesure. Par exemple, pour le capteur aéroporté AVIRIS, on dispose de 224 mesures spectrales, pour une résolution au sol de l'ordre de 20 mètres.

Dans ce travail, on s'intéresse à une analyse des textures présentes dans des images hyperspectrales tirant parti à la fois des relation spatiales et spectrales de celles-ci, visant ainsi à ajouter l'information spectrale aux méthodes classiques d'analyse de texture prenant en compte les variations spatiales des niveaux de gris.

Les méthodes d'analyse de texture couramment utilisées en traitement d'image, comme les matrices de coocurence [1] ou les filtres de Gabor [2], ne peuvent être adaptées simplement au cas des images multicomposantes qu'au prix d'une complexité algorithmique qui peut s'avérer rédhibitoire si l'on veut tirer parti des relations spectrales. Autrement, l'adaptation consiste à faire une analyse scalaire bande par bande, les unes indépendamment des autres, suivie par une étape nécessaire de fusion de données en classification. Pour mettre en œuvre l'analyse de texture telle qu'on la désire, une méthode vectorielle semble bien adaptée, car les méthodes de ce type prennent en compte les relations interbandes, qui, on le verra plus tard, sont porteuses d'information. Des approches de ce type ont déjà été considérées, en detection d'anomalies [3], par exemple.

L'application dans laquelle s'inscrit ce travail est la classification de zones urbaines. En effet, ces zones sont constituées de nombreux matériaux, ce qui est la source d'une grande variabilité de reflectance, qui se traduit à la fois entre différents pixels et à l'intérieur d'un même pixel dont la reflectance peut alors être un mélange de plusieurs composantes "pures". De plus, une partie de ces matériaux sont aussi caractéristiques de certains paysages naturels, donnant à certains pixels urbains les mêmes signatures que ces paysages naturels. De ce fait, l'étude des reponses radiométriques de ces zones ne peut pas aboutir à un résultat satisfaisant. Il est donc nécessaire d'étudier les zones urbaines à travers leur texture.

Lorsque l'on considère l'analyse de données de grande dimension par des méthodes statistiques, on est obligatoirement confronté aux problèmes liés au phénomène de Hughes [6], se caractérisant par la difficulté de réaliser de bonnes estimations dans des espaces qui se trouvent être pratiquement "vides". C'est la raison pour laquelle on effectue en général une projection de l'espace initial de façon à obtenir un espace de dimension plus réduite, facilitant ainsi l'estimation. Dans la première partie du rapport, on présentera plus en détails ce problème.

Dans la deuxième partie, on présente quelques méthodes classiques de réduction des données, leurs avantages et leurs limitations. On s'intéressera plus particulièrement à la poursuite de projection, une méthode de recherche de projection peu sensible au phénomène de Hughes et peu restrictive quant au choix du critère présidant à la projection. La poursuite de projection représente un cadre général de travail pour la recherche de sous-espace, qui a été adapté à la classification dans [7] sous l'appellation de poursuite de projection paramétrique. On propose une variante de cet algorithme à la fin de la seconde partie.

Dans une troisième partie, on propose une modélisation d'image hyperspectrale prenant en compte la texture, fondée sur les champs de Markov, ainsi qu'une extention de la poursuite de projection paramétrique adaptée à cette modélisation.

Enfin, on présente quelques résultats de classification obtenus par le couplage de la poursuite de projection et de la modélisation par champs de Markov. Ces résultats sont comparés à une approche par modélisation gaussienne des classes, ainsi qu'à des approches utilisant des méthodes classiques de réduction d'espace.

2 Nécessité et justification de la réduction de l'espace

2.1 Nécessité

2.1.1 Problèmes d'estimation

Lors d'un traitement statistique paramétrique, on passe en général par une ou plusieurs étapes d'estimation de paramètres. Dans le cas de la classification par exemple, il est nécessaire d'estimer les paramètres des lois de probabilité des différentes classes.

Or, la qualité de l'estimation de paramètres statistiques, étant donné un nombre d'échantillons d'apprentissage fixé, diminue avec le nombre de bandes car le nombre de paramètres à estimer (ie. d'inconnues) croît. Dans le cas de l'estimation de matrices de covariance, on peut même en arriver à obtenir des matrices singulières. Ceci se caractérise par une diminution du taux de bonne classification quand on augmente la dimension du vecteur observé tout en gardant constant le nombre de pixels d'apprentissage, c'est l'effet de Hughes [6].

2.1.2 Efficacité algorithmique

Une autre raison pour laquelle on désire réduire l'espace d'observation est d'ordre pratique : le volume de mémoire nécessaire au stockage et à la manipulation d'une image hyperspectrale est très grand, parfois jusqu'à 200 fois plus gros qu'une image monospectrale de même dimension. Par conséquent, la manipulation de ces images est très lourde : par exemple, estimer la variance d'une image monospectrale consiste en l'estimation d'une unique variable, alors qu'estimer la matrice de covariance d'une image à N bandes consiste en l'estimation de $\frac{N(N+1)}{2}$ variables. Or, la qualité d'un traitement par ordinateur dépend non seulement de la qualité du résultat obtenu, mais aussi de sa vitesse.

2.2 Justification

La réduction de la dimension de l'espace d'observation, à travers l'extraction d'un sous-espace de l'espace initial, si elle n'est pas trop pénalisante en terme de qualité du traitement envisagé, présente donc de gros avantages.

Or, plusieurs aspects caractéristiques des espaces de grande dimension en général, et des images hyperspectrales en particulier, font que cette réduction est possible au prix d'une perte d'information minimale.

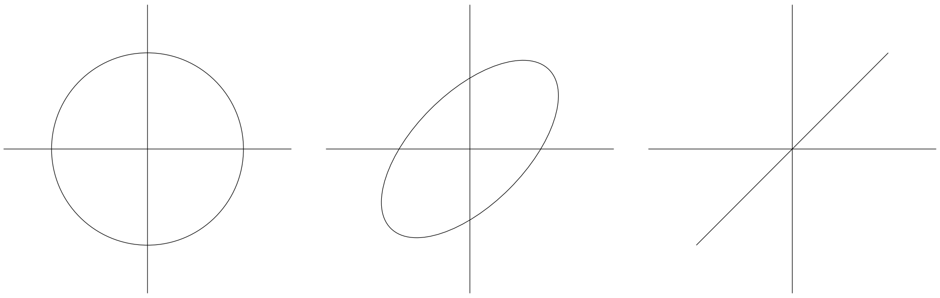
2.2.1 Renforcement de l'hypothèse de normalité

Une caractéristique des espaces de grande dimension est que les projections linéaires dans des sous-espaces ont tendance à rendre les distributions gaussiennes. Plus précisément, un jeu de données suivant une distribution quelconque dans un espace de dimension initiale N voit sa projection dans un sous-espace de dimension M fixée tendre vers un modèle gaussien quand N tend vers l'infini [8].

Ceci permet d'utiliser des algorithmes paramétriques fondés sur une hypothèse de normalité pour réaliser la classification. En particulier, si on utilise une classification par maximum de vraisemblance, les équations à résoudre sont très simples dans le cas gaussien.

2.2.2 Corrélation

La figure 2 représente la matrice de covariance inter-bande d'une scène AVIRIS. La matrice montre de fortes corrélations entre bandes adjacentes, par blocs. Les matrices de covariance de chaque classe présentent également la même structure très corrélée, par blocs. D'autre part, ces matrices présentent de grandes disparités eu égard à leurs valeurs propres. Ceci traduit une forte anisotropie des nuages considérés, ceux-ci sont donc très applatis et orientés (cf. figure 1), ce qui signifie qu'il existe une grande redondance entre les différents canaux, rendant inutile l'utilisation de l'observation dans toute sa dimension. La structure par blocs vient du fait que les propriétés physiques des matériaux sont communes par groupes de longueurs d'ondes proches, et changent radicalement pour certains paliers, comme par exemple les longueurs d'ondes d'absorption du rayonnement par certaines molécules de l'atmosphère (eau, oxygène...).



$\lambda_1 = 1,\ \lambda_2 = 1$ $\lambda_1 = 1,\ \lambda_2 = 0.5$ $\lambda_1 = 1,\ \lambda_2 = 0$

FIG. 1 – *Exemple en 2 dimensions de lignes iso-norme, traduisant les formes de nuages de points tirés selon des lois gaussiennes avec des matrices de covariance ayant différentes valeurs propres* λ_1 *et* λ_2

FIG. 2 – *Image de la matrice de covariance inter-bande d'une scène AVIRIS*

3 Méthodes de réduction d'espace

On a montré au chapitre précédent qu'il était intéressant d'utiliser pour le traitement des images hyperspectrales, non pas les mesures faites dans l'espace initial de grande dimension, mais une projection de celles-ci dans un espace de plus petite dimension. Le choix de ce sous-espace n'est, bien sûr, pas indifférent, et certains sous-espaces ont de bien meilleures caractéristiques que d'autres. Dans cette partie, on examine différents critères et différentes méthodes de projection.

3.1 Approches générales

Dans les méthodes suivantes, on considère un critère énergétique calculé sur l'image dans son ensemble, et non un critère relatif à la distinction des classes. Contrairement aux méthodes que l'on verra plus tard, elles ne sont pas optimales par rapport à des critères de classification, mais elles sont simples et donnent tout de même de bons résultats.

3.1.1 Analyse en composantes principales

Parmi les méthodes existantes, la méthode la plus simple, mais qui reste efficace, est l'analyse en composantes principales (ACP) [9, 10]. Le principe de cette méthode est de déterminer les «composantes principales» (CP) et de représenter le signal dans l'espace généré par les n premières CP, espace qui sera donc de dimension n. Les composantes principales sont un ensemble de vecteurs ayant 3 particularités intéressantes : premièrement, elles forment une base orthogonale de l'espace de départ ; deuxièmement, chaque composante principale est décorrelée de toutes les autres au regard du nuage de points total de l'image ; troisièmement, elles sont ordonnées (de façon décroissante) selon la variance que présente le nuage de points dans la direction des composantes.

Dans la base initiale, la matrice de covariance peut s'exprimer comme :

$$\Sigma = P^t \Lambda P \tag{1}$$

Λ étant la matrice de covariance dans la base après rotation des axes telle que :

$$\Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix} \tag{2}$$

avec les valeurs propres λ_i telles que $\lambda_1 > \lambda_2 > \cdots > \lambda_N$.

En prenant comme espace de représentation les n premières CP (qui correspondent aux n plus grandes valeurs propres), on est assuré d'avoir une approximation du signal telle que la moyenne de l'erreur sur l'image entière est nulle et sa variance minimale.

Cette méthode est optimale au sens de l'énergie du signal, c'est-à-dire que pour une valeur de n donnée, il est impossible de déterminer une base dans laquelle l'erreur quadratique moyenne entre le signal et son approximation soit inférieure. Pour n fixé, la projection sur les n premières CP est celle garantissant une perte d'énergie minimale. En revanche, elle a deux inconvénients majeurs. Le premier est le temps de calcul qu'elle nécéssite. Les composantes principales sont les vecteurs propres de la matrice de covariance *totale* de l'image (ie. qui doit être calculée sur la totalité de l'image) d'où la lenteur de cet algorithme. Ensuite, le calcul des vecteurs propres de la matrice peut être long, si le nombre de bandes N est très grand. Pour ce calcul, il est possible de profiter de la structure par blocs de la matrice pour diviser le problème en une somme de sous-problèmes plus simples [10].

Le choix de la dimension d'arrivée n peut être fait selon un critère énergétique, c'est-à-dire que si l'on choisit un rapport $\eta \in [0,1]$, on prendra n tel que dans la base d'arrivée :

$$\sum_{i=0}^{n} E\{y_i^2\} > \eta \sum_{i=0}^{N} E\{y_i^2\} \tag{3}$$

où y_i sont les coordonnées transformées et centrées en zero, et $E\{.\}$ représente l'espérance mathématique.

3.1.2 *Minimum Noise Fraction*

La transformation appelée *Minimum Noise Fraction* (MNF) est une variante de l'ACP, ajustée au bruit [11]. Cette transformation consiste en une rotation de la base afin d'obtenir une nouvelle base orthogonale où les vecteurs sont ordonnés suivant le rapport signal sur bruit (SNR), et non suivant la variance comme pour l'ACP.

Sa mise en œuvre pratique se décompose en deux ACP successives. La première ACP se fait par rapport aux statistiques du bruit, supposées connues. A l'issue de cette étape, on connait une base orthogonale dans laquelle le bruit est décorrélé. La base intermédiaire obtenue à l'issue de cette étape est la base de vecteurs propres de la matrice de covariance du bruit. Cette base est modifiée par une étape de normalisation consistant à diviser chaque vecteur par la racine de sa valeur propre associée, de sorte que la matrice de covariance du bruit dans la nouvelle base soit la matrice identité. On a *sphérisé* le bruit.

La matrice de covariance du bruit Σ_n peut s'ecrire comme suit :

$$\Sigma_n = P^{-1} I P \tag{4}$$

où la matrice de passage P est constituée de vecteurs orthogonaux mais n'ayant pas forcément une norme égale à 1, et I est la matrice identité.

Une fois dans cette base, on effectue la deuxième ACP, en considérant cette fois la matrice de covariance des données (incluant le bruit). On obtient une base de vecteurs propres, telle que la valeur propre associée à un vecteur de base est le SNR des données selon ce vecteur.

$$\Sigma_s = P^{-1} Q^t \Lambda Q P \tag{5}$$

où la matrice de passage Q est une matrice orthonormée, et Λ une matrice diagonale.

La matrice de passage totale pour aller de la base initiale à la base MNF est donc QP.

On est capable avec cette transformation de déterminer la dimension intrinsèque des données. En effet, si un vecteur de la base finale a pour valeur propre associée une valeur bien supérieure à 1, c'est que cette composante contient surtout de l'information et que sa variance est due à la variabilité propre des données. Tandis que si ce vecteur de base a pour valeur propre associée une valeur proche de 1, c'est que sa variance est principalement due au bruit. Ceci peut être une manière de déterminer une valeur de n adéquate, puisque les bandes contenant principalement du bruit ne sont pas intéressantes.

Le problème de cette méthode est qu'il faut avoir une connaissance *a priori* sur le bruit, ce qui n'est pas forcément possible si l'on ne connait pas les caractéristiques du capteur ou les conditions d'acquisition des données.

3.2 Approches pour la classification

Pour la classification, d'autres méthodes peuvent être envisagées, considérant non plus un critère énergétique comme l'ACP, mais étant fondées sur la distinction entre classes.

On se place, pour ces approches, dans le cadre de la classification supervisée car ces méthodes nécessitent une définition préalable des statistiques des classes. Or, du fait de l'effet de Hughes, calculer des statistiques devient moins trivial quand la dimension de l'observation devient de plus en plus grande.

Avec ces approches, il est possible de perdre beaucoup d'information, si celle-ci n'est pas un facteur de discrimination entre les classe définies. Parmi ces approches, deux méthodes nous ont paru intéressantes : l'*analyse discriminante* et la *poursuite de projection*.

3.2.1 Analyse discriminante

L'*analyse discriminante* [12] consiste en la maximisation d'un critère lié à la dispersion inter-classe et la dispersion intra-classe. La dispersion inter-classe est la dispersion des centres des classes que l'on doit maximiser, alors que la dispersion intra-classe est la dispersion d'une même classe autour de son centre, qui doit être minimisée.

On définit donc deux matrices, la matrice de dispersion intra-classe :

$$S_w = \sum_i P(\omega_i) S_i \tag{6}$$

et la matrice de dispersion inter-classe :

$$S_b = \sum_i P(\omega_i)(M_i - M_0)(M_i - M_0)^t \tag{7}$$

où $M_0 = \sum_i P(\omega_i) M_i$, et où M_i, S_i et $P(\omega_i)$ sont respectivement la moyenne, la matrice de covariance et la probabilité *a priori* de la classe ω_i.

On définit $S_w^d = d^t S_w d$ et $S_b^d = d^t S_b d$, comme étant les projections des matrices définies plus haut par une matrice de projection orthogonale d.

On peut construire différents critères à partir de ces matrices, comme ceux recensés dans [12]. En pratique, on utilise le plus souvent le critère suivant :

$$J = tr((S_w^d)^{-1} S_b^d) \tag{8}$$

qu'il faut maximiser par rapport à d.

L'optimisation conduit à un espace dont la dimension est limitée par le nombre de classes, car $S_w^{-1} S_b$ est de rang égal au nombre de classes moins 1.

D'autre part, si les classes ne sont pas assez uniformément réparties dans l'espace initial, les résultats de classification sont moins bons, notamment si deux classes ont des moyennes très proches. Malgré cela, la méthode est intéressante pour sa rapidité. Mais elle dépend fortement des statistiques calculées dans l'espace initial.

3.2.2 Poursuite de projection

La poursuite de projection [13] est une méthode visant à rechercher un sous-espace dans lequel un indice de projection est maximisé.

Cette méthode parait intéressante pour deux raisons :

- le phénomène de Hughes est atténué du fait que tous les calculs sont fait dans le sous-espace projeté,
- l'indice de projection peut être adapté selon le traitement visé.

Initialement [13], la poursuite de projection était utilisée dans des contextes non paramétriques, afin de trouver des angles de projection pour lesquels les données seraient "bien" représentées. Les indices de projection étaient alors des mesures liées à l'histogramme du nuage de points.

Plus récemment, cette méthode a été utilisée dans un contexte paramétrique dans le cadre de la classification d'image hyperspectrale [7]. Ayant choisi un modèle Gaussien non-contextuel, les auteurs proposent d'appliquer une approximation de la poursuite de projection afin de trouver le sous-espace maximisant la distance de Bhattacharyya entre les classes définies au préalable à travers des échantillons d'apprentissage.

De façon analogue, on propose d'utiliser la distance de Bhattacharyya dans l'application présente, car cette distance est reliée à une borne supérieure de la probabilité d'erreur [12, 14]. Néanmoins, cette distance sera calculée sur la probabilité conditionnelle définissant le modèle de texture considéré, que l'on verra plus tard.

La poursuite de projection est une approche globale du problème de la détermination de sous-espace vectoriel, applicable dans différents domaines.

L'approche est très générale puisqu'il s'agit de trouver un sous-espace par projection linéaire dans lequel une fonction appelée indice de projection est maximisée ou minimisée. Le choix de l'indice de projection est bien entendu le point crucial de l'algorithme, puisqu'un bon choix dépend presqu'intégralement de l'application. Par exemple, un choix d'indice pour la compression pourrait être la variance totale, ce qui ramène à faire une analyse en composantes principales.

Dans [7], une mise en œuvre approximée de la poursuite de projection dite séquentielle est décrite. On la rappelle dans la figure suivante (algorithme 1). L'ensemble des bandes est divisé en K sous-groupes de chacun desquels on extrait un axe de projection optimal estimé itérativement. Dans la mise en œuvre originale, les K groupes de bandes sont déterminés en divisant l'ensemble initial en groupes de même nombre de bandes. On utilise une variante de cet algorithme, dans laquelle les sous-groupes de bandes adjacentes n'ont pas forcément le même nombre de bandes, et sont choisis conformément à un critère que l'on décrit par la suite.

Algorithme 1 Poursuite de projection séquentielle

L'ensemble des bandes est divisé en K groupes de bandes adjacentes
$n \leftarrow 0$, $arret \leftarrow 0$
pour $i = k$ à K **faire**
  Initialiser arbitrairement $v_k = v_k^0$
fin pour
tant que $arret \neq 1$ **faire**
  pour $k = 1$ à K **faire**
    Soit $v(x) = (v_1^n, \ldots, v_{k-1}^n, x, v_{k+1}^{n-1}, \ldots, v_K^{n-1})$, avec x combinaison linéaire des vecteurs du k^{eme} groupe de vecteurs.
    $v_i^n \leftarrow \arg\max_x I(x)$, $I(x)$ étant un indice de projection à définir
  fin pour
  si $\|v^n - v^{n+1}\| < \epsilon$ **alors**
    $arret \leftarrow 1$
  fin si
  $n \leftarrow n + 1$
fin tant que

Poursuite de projection séquentielle avec groupes de bandes variables

Nous proposons dans ce rapport une variante de la poursuite de projection séquentielle, pour laquelle on cherche un découpage du spectre en groupes de bandes plus adapté que le découpage à pas constant.

Etant donné que chaque groupe est utilisé pour extraire un axe de projection, un critère pour le découpage peut être l'homogénéité de la répartition de l'indice de projection dans les différents groupes.

Soit $z = \{z_0, \ldots, z_K\}$ les bornes définissant un découpage du spectre en K sous-ensembles de bandes, avec $z_0 = 0$ et $z_K = N$, N étant le nombre total de bandes. Un groupe de bande k, duquel on va extraire le vecteur v_k de la base d'arrivée, sera donc constitué des bandes de l'espace original comprises entre les borne z_i et z_{i+1} (z_i inclus).

On cherche à trouver le découpage $\hat{z}$ tel que :

$$\hat{z} = \arg\min_z \sigma_I(z) \tag{9}$$

avec

$$\sigma_I(z) = \frac{1}{K} \sum_{i=0}^{K-1} (I(z_i, z_{i+1}) - m_I(z))^2 \tag{10}$$

où $I(z_i, z_{i+1})$ est la valeur de l'indice correspondant à une projection sur l'espace généré par le groupe des bandes comprises entre z_i et z_{i+1} (z_i inclus), et

$$m_I(z) = \frac{1}{K} \sum_{i=0}^{K-1} I(z_i, z_{i+1}) \tag{11}$$

Or

$$\sigma_I(z) = \frac{1}{K} \sum_{i=0}^{K-1} I^2(z_i, z_{i+1}) - m_I^2(z) \tag{12}$$

Pour minimiser σ_I, et trouver une estimation de $\hat{z}$, on examine les sites z_i, $i = 1, \ldots, K-1$, un par un, et on leur assigne, à chaque passage, la valeur qui minimise l'énergie, les autres bornes z_j, $j \neq i$, étant fixées (qui est une fonction convexe de z_i), cette valeur étant comprise entre les z_{i-1} et z_{i+1} courants. Ceci est illustré par la figure 3. L'algorithme est itéré jusqu'à obtenir la convergence.

Cet algorithme ne garantit pas de trouver le minimum global d'énergie, mais il est plus rapide que les algorithmes d'optimisation globale et aboutit à de bons résultats

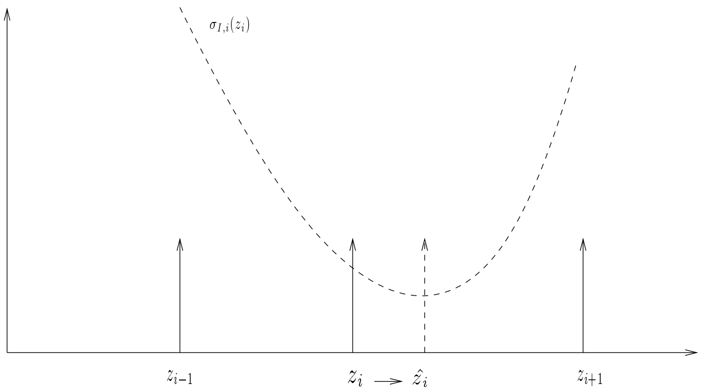


FIG. 3 – *Illustration de l'algorithme d'ajustement des bornes : pour z_{i-1} et z_{i+1} fixés, la valeur optimale $\hat{z}_i$ de z_i est celle minimisant la fonction $\sigma_I(z_i)$*

si l'initialisation est bonne. Pour celle-ci, on choisit z tel que l'on ait un pas constant entre les z_i, c'est à dire que l'on considère le découpage utilisé dans [7].

La figure 4 montre les évolutions comparées de la distance optimale estimée en fonction du nombre de bandes dans le cas de deux classes présentes dans des données AVIRIS, pour les deux mises en œuvre de la poursuite de projection séquentielle. Il apparaît (dans ce cas représentatif) qu'il n'est pas toujours bénéfique d'utiliser cet ajustement du découpage quand l'espace final est de dimension supérieure à 20. En revanche, quand on considère des dimensions finales inférieures à 20, l'indice de projection optimal est supérieur si l'on opère un ajustement des groupes de bandes.

Le problème de la détermination d'un bon indice de projection reste ouvert, on en discutera plus loin dans ce rapport.

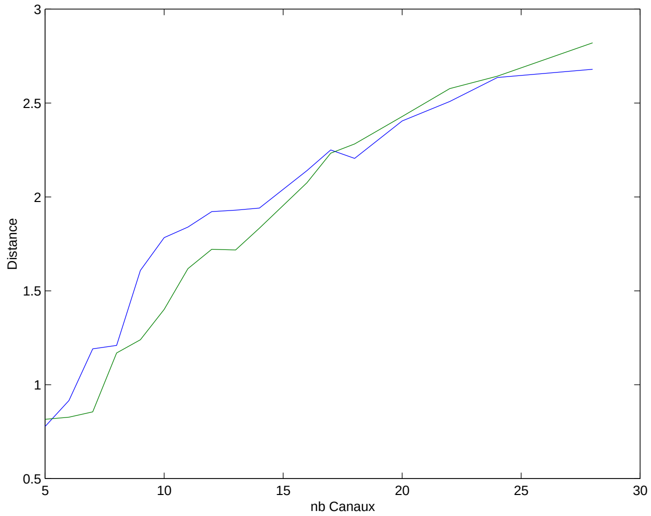


FIG. 4 – *Distance optimale estimée en fonction du nombre de bandes pour les deux mises en œuvre de la poursuite de projection séquentielle, dans le cas de deux textures d'une même scène AVIRIS. en bleu : mise en œuvre avec groupes variables, en vert : mise en œuvre avec groupes fixes.*

4 Modélisation

Les approches en matière d'analyse de texture sont nombreuses et il serait difficile d'en faire une revue exhaustive. Parmi les plus populaires, et sans doutes les plus efficaces en analyse de texture monodimentionnelle, on trouve les approches utilisant les matrices de cooccurence pour extraire des paramètres caractéristiques de la texture observée [1]. Cette approche donne de bons résultats pour les images monodimentionnelles, mais son adaptation aux images multidimentionnelles parait irréalisable. En effet, une matrice de coocurence est une matrice de dimention $N \times N$, où N est le nombre de niveaux de gris quantifiés de l'image. Pour adapter cette méthode aux images hyperspectrales, il faut soit travailler bande par bande de façon scalaire, auquel cas l'information inter-bande est perdue, soit travailler sur le cube image. Mais ce dernier cas est pratiquement impossible, car il impliquerait de travailler sur des matrices gigantesques $N_q N_b \times N_q N_b$, où N_q est le nombre de niveaux de gris, et N_b le nombre de bande. Or les images AVIRIS ont 224 bandes et sont codées sur 16 bits!

L'approche par filtrage de Gabor [2] est plus rapide, mais l'extension aux données multi-variées sans perte d'information n'est pas évidente, pour les mêmes raisons que précédemment concernant la prise en compte de l'information inter-bande.

La modélisation par champs de Markov, que l'on présente dans les paragraphes suivants, est intéressante du fait que l'on travaille sur des champs de variables aléatoires qui peuvent être des vecteurs aléatoires. Or, le fait d'avoir une méthode d'analyse fondée sur une modélisation vectorielle, donc travaillant sur l'ensemble des bandes conjointement, permet de prendre en compte les relations inter-bandes.

4.1 Modèles markoviens

Les champs de Markov sont des outils très utilisés en traitement d'image. Leur intérêt est principalement de pouvoir modéliser des interactions entre pixels de façon simple, à travers un modèle probabiliste. Un champ de Markov X est un champ aléatoire présentant la propriété suivante :

$$P(X_s = x_s / X^s = x^s) = P(X_s = x_s / \{x_t\}, t \in V_s) \tag{13}$$

Ceci signifie que la probabilité en un site s conditionnellement au reste du champ est égale à la probabilité en ce site connaissant simplement ses voisins. Cette modélisation permet d'introduire explicitement des interactions à courte échelle, et induit implicitement des interactions à plus longue échelle par "transitivité".

C'est ainsi qu'ils sont efficacement utilisés en segmentation/classification ou en restauration, pour régulariser les images d'étiquettes ou de niveaux de gris.

Mais, outre la régularisation, les champs de Markov peuvent aussi être utilisés pour la modélisation des images, en particulier, des textures fines [15, 16]. Dans ce cadre, il existe de nombreux modèles markoviens de texture, comme par exemple le modèle auto-binomial [17], le modèle de Derin-Elliott [18] ou le modèle markovien gaussien aussi appelé auto-normal, que l'on détaille par la suite. Bien qu'il existe des modèles markoviens de textures plus sophistiqués, le modèle markovien gaussien est particulièrement bien adapté au problème d'analyse de textures urbaines, comme indiqué ci-dessous.

4.2 Modèle markovien gaussien

Le modèle markovien gaussien a beaucoup été utilisé en imagerie mono-dimensionnelle pour l'analyse de texture, notamment urbaine [19, 20]. Mais il est aussi possible d'utiliser cette modélisation dans le cadre de l'imagerie multi-spectrale ; deux approches peuvent être distinguées dans ce cas. Dans la première approche [3, 21], les données couleur ou hyperspectrales sont considérées comme une collection de variables scalaires dans un modèle à trois dimensions : deux dimensions spatiales et une dimension spectrale. La deuxième approche, utilisée notamment dans [22], consiste à définir un modèle à deux dimensions sur les données vectorielles que constituent les observations aux différentes longueurs d'onde d'un même site. Dans cette deuxième approche, les interactions entre les valeurs à différentes longueurs d'onde ne sont pas modélisées explicitement, mais elles existent tout de même, et découlent des paramètres qui sont estimés à partir des données. Il n'est donc pas nécessaire de faire d'hypothèse concernant la portée de ces interactions.

Les méthodes présentées dans ce rapport sont reliées à une modélisation de ce dernier type, que l'on appelle MGMRF (*Multivariate Gaussian Markov Random Field*).

4.3 Le MGMRF

4.3.1 Loi conditionnelle

Soit une image x de dimensions spatiales N_hxN_v et ayant un nombre de bandes N_b. Ce champ est supposé être une réalisation du champ X. Comme on l'a vu au paragraphe précédent, on suppose un champ à 2 dimensions (les dimensions spatiales) composé de variables vectorielles de dimension égale au nombre de bandes (les réponses spectrales), la modélisation markovienne implique que la probabilité d'un vecteur en un site s connaissant le reste de l'image est égale à la probabilité en ce site connaissant ses voisins. La modélisation markovienne gaussienne ajoute une contrainte sur cette probabilité, qui est supposée être de la forme :

$$P(x_s|\{x_t\},t \in V_s) = \frac{1}{Z}\exp\left\{-\frac{1}{2}\|x_s - \sum_{t \in V_s}\theta_{t-s}x_t\|^2_\Sigma\right\} \tag{14}$$

Avec $\|a\|^2_\Sigma = a^t\Sigma^{-1}a$, les θ_i étant des matrices de transformation et Σ la matrice de covariance conditionnelle.

Cette loi est donc une loi gaussienne dont la covariance est constante et la moyenne dépendante du voisinage.

Cette expression est valable si l'on considère des données centrées, c'est-à-dire de moyenne nulle. Dans le cas de moyennes non nulles, l'expression est légèrement modifiée :

$$P(x_s|\{x_t\},t \in V_s) = \frac{1}{Z}\exp\left\{-\frac{1}{2}\|x_s - \sum_{t \in V_s}\theta_{t-s}x_t - (Id - \sum_{t \in V_s}\theta_{t-s})\mu\|^2_\Sigma\right\} \tag{15}$$

Par la suite, on considèrera que la texture homogène considérée a été centrée, la généralisation étant automatique.

La formulation (14) permet d'écrire :

$$x_s = \sum_{t \in V_s}\theta_{t-s}x_t + u_s \tag{16}$$

avec u_s bruit blanc gaussien, centré, de matrice de covariance Σ.

De plus, pour des contraintes de cohérence, on doit avoir $\theta_i = \theta_{-i}$, ce qui conduit à :

$$x_s = \sum_{k=0}^{K}\theta'_k m_s^k + u_s \tag{17}$$

où k désigne chaque type de voisinage (horizontal, vertical,...), K le nombre de types de voisins, m_s^k la moyenne des voisins de type k, et $\theta'_k = \theta_k n_k$, avec n_k le nombre de voisins de type k. Un *type* de voisins désigne, ici, l'ensemble des voisins t d'un pixel s ayant la même contribution dans le modèle, ce qui se traduit par par le même paramtètre θ_{s-t}. A titre d'exemple, dans un système de voisinage en 4-connexité isotrope, les 4 voisins sont considérés comme étant du même type, tandis que dans un système voisinage en 8-connexité anisotrope, on peut considérer quatre types de voisins : les voisins horizontaux, verticaux, et les deux types de voisins correspondant aux deux diagonales.

4.3.2 Loi globale

Pour obtenir la formule de la loi globale, on revient sur la formule (16).

$$u_s = x_s - \sum_{t \in V_s} \theta_{t-s} x_t \tag{18}$$

Soit x l'image entière réorganisée en un vecteur colonne, $x = (x^t_{(0,0)} \dots x^t_{(N_v,N_h)})^t$, de même $u = (u^t_{(0,0)} \dots u^t_{(N_v,N_h)})^t$.

$$u = Ax \tag{19}$$

Le vecteur u suit une loi normale, donc le vecteur x également. On peut écrire :

$$P(x) = \frac{1}{Z(S)} \exp\left(-\frac{1}{2} x^t S^{-1} x\right) \tag{20}$$

avec $S = E\{xx^t\} = A^{-1} E\{ux^t\}$, et $Z(S)$ constante de normalisation.
Or, d'après le principe d'orthogonalité, on doit avoir :

$$E\{u_i x^t_j\} = \delta_{i,j} \Sigma, \ \forall i,j \tag{21}$$

Cela se traduit sur la matrice globale de la façon suivante :

$$E\{ux^t\} = I_{N_h} \otimes I_{N_v} \otimes \Sigma \tag{22}$$

où $\otimes$ désigne le produit de Kronecker [1] [23], et I_n la matrice identité de dimension n.

Donc l'équation (20) peut être réecrite:

$$P(x) = \frac{1}{Z(S)} \exp\left(-\frac{1}{2} x^t \left(I_{N_h} \otimes I_{N_v} \otimes \Sigma^{-1}\right) Ax\right) \tag{23}$$

Ce qui permet de déduire directement l'expression des potentiels de la formulation de Gibbs :

$$P(x) = \frac{1}{Z} \exp\left(-\sum_s V_0(x_s) - \sum_{(s,t) \in C_1} V_1(x_s,x_t)\right) \tag{24}$$

1. le produit de kronecker entre deux matrices $A = \{a_{ij}\}$ de dimension n et $B = \{B_{ij}\}$ de dimension m est une matrice $A \otimes B = C = \{c_{ij}\}$ de dimension mn telle que $c_{im+k,jm+l} = a_{i,j} b_{k,l}$

$$V_0(x_s) = \frac{1}{2} x_s^t \Sigma^{-1} x_s \tag{25}$$

$$V_1(x_s,x_t) = -\frac{1}{2}\left(x_s^t \Sigma^{-1} \theta_{t-s} x_t + x_t^t \Sigma^{-1} \theta_{t-s} x_s\right) \tag{26}$$

4.3.3 Caractéristiques des espaces de grande dimension

Après avoir présenté ce modèle dans sa forme générale, on s'intéresse plus particulièrement à son application dans le cadre de la classification des images AVIRIS à 224 bandes sprectrales.

Le processus de classification tel qu'on l'envisage est constitué dans un premier temps d'une phase d'apprentissage durant laquelle les classes à distinguer sont définies par un opérateur, puis caractérisées par l'estimation de leurs paramètres.

Le nombre de paramètres du modèle markovien gaussien est variable, selon l'ordre du voisinage, son anisotropie, ou les contraintes appliquées à la matrice θ. Dans le cas par exemple d'un modèle anisotrope à 4 voisins, il est nécessaire de calculer les deux matrices θ_v et θ_h correspondant aux voisinages vertical et horizontal, ainsi que la matrice de covariance conditionnelle Σ. Soit, si le vecteur d'observation est de dimension N, $3N^2$ paramètres. Dans le cas de données AVIRIS à 224 bandes, l'estimation d'un nombre aussi élevé de paramètres est très difficile, surtout étant donné le nombre d'échantillons disponibles pour effectuer le calcul, qui ne peut excéder quelques milliers. Or plus le nombre de paramètres est élevé, plus le phénomène de Hughes [6] est présent.

Deux solutions peuvent être apportées à ce problème. D'une part, comme mentionné au début de ce rapport, il est possible de choisir des critères et des méthodes permettant de réduire la dimension des observations de façon à améliorer estimation et classification. D'autre part, il est possible de simplifier le modèle de façon à ce que le nombre de paramètres à estimer soit radicalement diminué.

Dans le cas de cette application, la simplification consistera à donner une certaine forme *a priori* aux matrices de transformation θ_i :

$$\theta_i = \mathrm{diag}(a_i) = a_i \ \mathrm{Id} \tag{27}$$

Ceci consiste à considérer que l'on à d'une part indépendance entre les bandes dans les interactions entre pixels, et d'autre part que ces interactions sont identiques dans les différentes longueurs d'onde du spectre. Cett hypothèse a pour conséquence sur la probabilité conditionnelle de lui faire prendre une forme simple :

$$P(x_s|\{x_t\},t \in V_s) = \frac{1}{Z} \exp\left\{ -\frac{1}{2} \|x_s - \sum_{t \in V_s} a_{t-s} x_t\|_{\Sigma}^2 \right\} \tag{28}$$

C'est-à dire que les potentiels ont la forme suivante :

$$V_0(x_s) = \frac{1}{2} x_s^t \Sigma^{-1} x_s \tag{29}$$

$$V_1(x_s,x_t) = -\frac{a_{t-s}}{2} \left(x_s^t \Sigma^{-1} x_t \right) \tag{30}$$

Dans ce cas, la probabilité conditionnelle s'écrit :

$$P(x_s|\{x_t\},t \in V_s) = \frac{1}{Z'} \exp\left\{ -\sum_{t \in V_s} \alpha_{t-s} \|x_s - x_t\|_{\Sigma}^2 - \beta \|x_s\|_{\Sigma}^2 \right\} \tag{31}$$

4.3.4 Domaine de validité des paramètres

Pour que le champ X de loi (20) existe, il faut que la matrice S soit définie positive, par conséquent il faut que A soit définie positive si l'on considère que Σ l'est. Cette dernière supposition est vraisemblable, si l'on considère un nombre d'échantillons suffisamment élevé.

Par la suite, on prendra la formulation de la probabilité conditionnelle suivante :

$$P(x_s|\{m_k\}) = \frac{1}{Z} \exp\left\{ -\frac{1}{2} \|x_s - \sum_{k} a'_k m_s^k\|_{\Sigma}^2 \right\} \tag{32}$$

où les a'_k sont obtenus à partir des a_k comme les θ' à partir des θ.

Par un calcul similaire à celui mené dans [22], on montre qu'une condition nécessaire pour l'existence du champ markovien gaussien de paramètres $\{a'_k\}$ est que $\sum_{k=0}^{K-1} a'_k < 1$, où K désigne le nombre de voisins.

4.3.5 Estimation des paramètres

Considérons une image x de taille $N_v \times N_h$ de moyenne nulle, supposée avoir été générée par le modèle décrit par l'équation (28)dans ce chapitre, dont les paramètres sont inconnus. On désire estimer ces paramètres d'une façon rapide et précise.

Un critère très utilisé pour l'estimation de paramètres est le critère du Maximum de Vraisemblance (MV). Il consiste à retenir le vecteur de paramètres ϕ (dans la cas présent $\phi = (\{a_k\},\Sigma)$), tel que :

$$\phi = \arg\max_{\phi}(P(x/\phi)) \tag{33}$$

En utilisant l'équation (20) et en explicitant la constante de normalisation, il vient :

$$P(x/\phi) = \frac{(2\pi)^{-\frac{N_v N_h N_b}{2}}}{|S|} \exp\left(-\frac{1}{2}x^t S^{-1} x\right) \tag{34}$$

La log-vraisemblance $LV(\phi)$ s'écrit donc :

$$LV(\phi) = C - \ln|S| - \frac{1}{2}x^t S^{-1} x \tag{35}$$

La maximisation de cette expression est complexe, du fait du calcul du déterminant de S, même si cela reste possible pour certaines variantes du modèle [3, 24].

Un critère permettant une estimation plus rapide est celui du maximum de pseudo-vraisemblance [17]. Ce critère est fondé sur une hypothèse d'indépendance des probabilités conditionnelles et mène à l'approximation suivante concernant la probabilité du champ x :

$$P(x) \approx \prod_{s\in\Omega} P(x_s|\{x_t\}, t \in V_s) \tag{36}$$

L'estimateur de maximum de pseudo-vraisemblance est :

$$\phi = \arg\max_{\phi}\left(\prod_{s\in\Omega} P(x_s|\{x_t\},\phi)\right) \tag{37}$$

$$\phi = \arg\min_{\phi}\left(\sum_s \|x_s - \sum_k a'_k m^k_s)\|^2_{\Sigma} + N_v N_h \ln|\Sigma|\right) \tag{38}$$

Une fois de plus, la minimisation ne se ramène pas à une formule simple, notamment parce que la matrice Σ dépend des $\{a'_k\}$. Ce qui signifie que la minimisation doit se faire sur les $\{a'_k\}$. En considérant la matrice Σ indépendante de ces coefficients, la minimisation selon les $\{a'_k\}$ est ramenée à un calcul matriciel :

$$\begin{pmatrix} \langle (m^0_s)^t \Sigma^{-1} m^0_s \rangle & \dots & \langle (m^0_s)^t \Sigma^{-1} m^K_s \rangle \\ \vdots & & \vdots \\ \langle (m^K_s)^t \Sigma^{-1} m^0_s \rangle & \dots & \langle (m^K_s)^t \Sigma^{-1} m^K_s \rangle \end{pmatrix} \begin{pmatrix} a'_0 \\ \vdots \\ a'_K \end{pmatrix} = \begin{pmatrix} \langle x^t_s \Sigma^{-1} m^0_s \rangle \\ \vdots \\ \langle x^t_s \Sigma^{-1} m^K_s \rangle \end{pmatrix} \tag{39}$$

On approxime alors la matrice de covariance conditionnelle à la matrice identité (ce qui ramène le problème à une simple estimation aux moindres carrés) ou à la matrice de covariance générale, de façon à pouvoir calculer les paramètres en deux étapes rapides : d'abord les $\{a'_k\}$, ensuite Σ.

Une fois les $\{a'_k\}$ calculés, le calcul de Σ est un simple calcul de covariance :

$$\Sigma = \langle u_s u_s^t \rangle = \left\langle (x_s - \sum_k a'_k m_s^k)(x_s - \sum_k a'_k m_s^k)^t \right\rangle \tag{40}$$

5 Application de la modélisation proposée à la réduction d'espace pour la classification

5.1 Mise en relation de la règle de classification et de l'indice de projection

Soit un problème de classification à deux classes, C_0 et C_1. La règle de décision correspondant au maximum de vraisemblance (MV) est, dans le cas de la modélisation présentée :

$$P_0(x_s/\{m_s^k\}) \underset{h_1}{\overset{h_0}{\gtrless}} P_1(x_s/\{m_s^k\}) \tag{41}$$

Dans la modélisation que l'on propose, les probabilités conditionnelles sont gaussiennes. Or, pour ces distributions, la distance de Bhattacharyya a une expression exacte simple. On peut donc définir une distance de Bhattacharyya conditionnellement aux $\{m_s^k\}$. Pour simplifier l'écriture, considérons un modèle à un seul type de voisinage, par exemple un voisinage 4-connexe isotrope. On a donc :

$$x_s \sim \mathcal{N}(a'_i m_s + b_i, \Sigma_i) \tag{42}$$

si x_s appartient à la classe C_i, m_s étant la moyenne des 4 voisins de s.

L'apparition du vecteur $b_i = (1 - a'_i)\mu_i$, où μ_i est la moyenne de la texture i, vient du fait que l'on ne peut plus considérer les données comme centrées, car cela reviendrait à se restreindre au cas très particulier où toutes les textures ont la même moyenne. On revient donc à la formule pour les données non centrées. Cela ne change en rien le processus d'estimation. En effet, on peut calculer la moyenne de la classe considérée et ainsi se ramener au cas de données centrées sur celle-ci.

La distance de Bhattacharyya conditionnelle entre les classes C_0 et C_1 s'écrit donc :

$$d_B(m_s) = \frac{1}{8}\|\mu_0(m_s) - \mu_1(m_s)\|_{\Sigma_{01}} + \frac{1}{2}\ln\left(\frac{|\Sigma_{01}|}{\sqrt{|\Sigma_0|\,|\Sigma_1|}}\right) \tag{43}$$

avec $\mu_i(m_s) = a_i' m_s + b_i$ et $\Sigma_{01} = \frac{\Sigma_0+\Sigma_1}{2}$.

Etant donné la règle de décision (41), la probabilité d'erreur connaissant m_s est liée à $d_B(m_s)$. En maximisant $d_B(m_s)$, on minimise cette probabilité d'erreur. On peut construire deux indices de projection reliés à cette distance de Bhattacharyya. La mise en œuvre de la poursuite de projection consistera à rechercher la projection dans laquelle l'un ou l'autre de ces indices sera maximisé.

Indice 1

$$I_1 = E\{b(m_s)\} = \int d_B(m_s) f(m_s) dm_s \tag{44}$$

Indice 2

$$I_2 = \min_{m_s} d_B(m_s) \tag{45}$$

Le problème de l'indice 1 est que l'on n'a accès à la loi $f(x)$ qu'à travers l'histogramme de l'image, ce qui peut augmenter le temps nécessaire à l'intégration. Par ailleurs, l'indice 2 à l'avantage de rester relié à une borne supérieure de la probabilité d'erreur ce qui lui donne un avantage en plus du temps de calcul. Mais, par ailleurs, l'indice 1 moyennant les contributions, évite de prendre en compte les contributions de m_s peu représentatives. Le choix entre les deux indices n'est donc pas trivial. Toutefois, de façon empirique, on remarque que les résultats de clasification obtenus lorsque l'on utilise l'indice 1 pour la réduction d'espace sont en général meilleurs que lorsque l'on utilise l'indice 2. Les résultats présentés par la suite sont donc obtenus avec le premier indice.

5.2 Problème à plusieurs classes

Le cas où le nombre de classes est supérieur à 2 est beaucoup plus vraisemblable dans le cadre de la télédétection. Or, la distance de Bhattacharyya ne peut se calculer qu'entre 2 distributions. Une méthode pour construire l'indice correspondant à un problème de ce type consiste à prendre la plus petite des distances entre classes :

$$I = \min_{i,j} I_{i,j} \tag{46}$$

$I_{i,j}$ représentant un des deux indices proposés au sous-chapitre 5.1 correspondant au problème de classification entre i et j.

On peut même aller plus loin, puisque si l'on est intéressé par la classification des zones urbaines, il suffit de ne chercher ce minimum que dans l'espace des couples de classes dont l'une d'elle au moins est une classe urbaine. C'est-à-dire que les distances entre classes non urbaines ne nous intéressent pas, car on ne cherche pas a les classifier. Par conséquent, il n'est pas nécessaire de faire en sorte que la distance entre ces classes soit grande.

5.3 Calcul de l'indice de projection

Etant donné un sous-espace $E' \subset E$, le calcul de l'indice n'est pas immédiat. Il faut, en effet, procéder en deux étapes. La première consiste à réestimer les paramètres du modèle, car un des intérêts de la méthode est d'effectuer les calculs dans le sous-espace. La deuxième est la projection des réponses et des matrices dans le sous-espace et le calcul de l'indice.

L'estimation des $\{a'_k\}$ se fait par l'équation (39), qui, si l'on approxime la matrice de covariance conditionnelle par la matrice identité, peut se réécrire comme suit :

$$\begin{pmatrix} \mathrm{tr}\left\langle m_s^0 (m_s^0)^t \right\rangle & \dots & \mathrm{tr}\left\langle m_s^K (m_s^0)^t \right\rangle \\ \vdots & & \vdots \\ \mathrm{tr}\left\langle m_s^0 (m_s^K)^t \right\rangle & \dots & \mathrm{tr}\left\langle m_s^K (m_s^K)^t \right\rangle \end{pmatrix} \begin{pmatrix} a'_0 \\ \vdots \\ a'_K \end{pmatrix} = \begin{pmatrix} \mathrm{tr}\left\langle m_s^0 x_s^t \right\rangle \\ \vdots \\ \mathrm{tr}\left\langle m_s^K x_s^t \right\rangle \end{pmatrix} \tag{47}$$

Soit P une matrice de projection, l'estimation des $\{a'_k\}$ dans l'espace image du projecteur correspond à la résolution de l'équation suivante :

$$\begin{pmatrix} \mathrm{tr}(P\Sigma_{00}P^t) & \dots & \mathrm{tr}(P\Sigma_{0K}P^t) \\ \vdots & & \vdots \\ \mathrm{tr}(P\Sigma_{0K}P^t) & \dots & \mathrm{tr}(P\Sigma_{KK}P^t) \end{pmatrix} \begin{pmatrix} a'_0 \\ \vdots \\ a'_K \end{pmatrix} = \begin{pmatrix} \mathrm{tr}(P\Sigma_{x,0}P^t) \\ \vdots \\ \mathrm{tr}(P\Sigma_{x,K}P^t) \end{pmatrix} \tag{48}$$

avec $\Sigma_{ij} = \left\langle m_s^i (m_s^j)^t \right\rangle$ et $\Sigma_{x,i} = \left\langle m_s^i x_s^t \right\rangle$.

A la suite du calcul des $\{a'_k\}$, on peut réestimer Σ. On a alors calculé tous les paramètres du modèle nécessaires au calcul de la distance de Bhattacharyya.

La deuxième étape du calcul de l'indice consiste à calculer les valeurs de la distance de Bhattacharyya, connaissant les paramètres, correspondant aux différentes valeurs de m_s, pour évaluer l'indice de projection.

FIG. 5 – *Affichage en fausse couleur des canaux {30,20,10} d'un collage de textures extraites d'une scène AVIRIS prise au-dessus de Moffett Field en Californie.*

6 Résultats

6.1 Collage de textures urbaines

Avant de voir les résultats sur des scènes réelles, ces méthodes sont étudiées sur un collage de différentes textures extraites d'une image AVIRIS prise en 1997 au-dessus de la région de Moffett Field, en Californie. Les résultats sur ces collages sont plus quantifiables, puisque l'on connait la vérité de terrain. La figure 5 représente un affichage des trois canaux 30, 20 et 10 de ce collage. L'image est constituée de quatre textures différentes de petite dimension 64x64, représentant des tissus urbains différents. Les échantillons utilisés pour l'apprentissage sont les pixels du carré central de chaque sous-image, de dimension 11x11, soit 121 pixels par texture. Les résultats montrent le pouvoir discriminant des méthodes testées, dans le cas de variabilité faible des textures, et pour des ensembles d'apprentissage réduits.

On compare les résultats de classification donnés par plusieurs méthodes et diverses modélisations dans les tableaux 6 et 7. Le tableau 6 montre les résultats lorsque l'espace réduit est de dimension 10, et le tableau 7 les résultats pour un espace projeté de dimension 20. L'analyse en composantes principales est notée ACP, elle est suivie d'une classification par maximum de vraisemblance gaussien pixélique. La poursuite de projection fondée sur le même modèle gaussien pixélique est notée PP-G, et le critère de classification est aussi le maximum de vraisemblance associé au modèle gaussien. La poursuite de projection utilisant un modèle markovien gaussien est notée PP-MG, et la classification associée repose sur un critère de maximum de vraisemblance associé au modèle markovien gaussien.

Type de réduction d'espace	Taux de classification (%)
ACP	84.68
PP-G	84.32
PP-MG	82.16

FIG. 6 – *Taux de bonne classification du collage de textures par différents modèles et méthodes de réduction d'espace (le sous-espace extrait est de dimension 10)*

Type de réduction d'espace	Taux de classification (%)
ACP	84.02
PP-G	82.9
PP-MG	79.8

FIG. 7 – *Taux de bonne classification du collage de texture par différents modèles et méthodes de réduction d'espace (le sous-espace extrait est de dimension 20)*

On constate que, dans le cas de ces collages d'image urbaine, utiliser une classification fondée sur la texture n'apporte rien, et est même préjudiciable aux résultats, même si le taux de bonne classification dans le cas du PP-MG est altéré par les transitions très abruptes entre textures, qui créent des effets de bords pour les pixels limitrophes, dont certains voisins n'appartiennent pas à la même texture.

Par conséquent, pour des textures à faible variabilité, les modèles pixéliques sont capables de saisir la plupart de l'information présente. D'autre part, on constate aussi que le phénomène de Hughes est déjà présent quand on utilise 20 bandes pour 121 échantillons d'apprentissage, puisque le taux de bonne classification est plus faible qu'avec 10 bandes.

6.2 Résultats sur une scène AVIRIS

On effectue, dans un deuxième temps, des tests de classification sur l'image réelle AVIRIS de Moffett Field.

Dans ces tests, on compare les résultats de classification utilisant la poursuite de projection fondée sur deux modélisations différentes, la première étant la modélisation gaussienne utilisée par Jimenez et Landgrebe dans [7], et la seconde étant la modélisation proposée dans ce rapport. Les résultats sont régularisés par un modèle de Potts [25].

FIG. 8 – *Affichage en fausse couleur des canaux {30,20,10} d'une scène AVIRIS prise au-dessus de Moffett Field en Californie.*

La figure 8 montre un affichage en fausses couleurs de 3 bandes du spectre visible de l'image de test. L'image est approximativement divisée d'une part en une zone de végétation et d'eau, et d'autre part en une zone urbaine. La zone urbaine, que nous considérons avec plus d'attention, est elle-même divisée en deux zones : une zone de petits bâtiments (des habitations), et une zone de grands bâtiments assez larges, à toits plus réfléchissants.

Dans toutes les figures, les résultats montrent la classification des zones urbaines, les autres classes étant regroupées en une classe globale dite de "non-urbain".

La figure 9 montre trois résultats de classification de la même image. Le premier résultat est celui donné par la méthode proposée en gardant 10 bandes, le deuxième est celui donné par la méthode de Jimenez et Landgrebe [7] par poursuite de projection sur modèle gaussien pixélique en gardant également 10 bandes, et le dernier résultat est celui de la classification par maximum de vraisemblance sur un modèle gaussien sur les 10 premières composantes principales.

En ce qui concerne la détection du masque urbain, il semble que les trois méthodes aient à peu de chose près la même efficacité, bien que les deux dernières méthodes produisent des fausses alarmes au niveau des eaux peu profondes de l'étang. En dehors de ces fausses alarmes, et au niveau de l'analyse intra-urbaine, on note une efficacité équivalente des deux méthodes de poursuite de projection comparées à l'ACP qui a tendance à détecter de gros bâtiments dans les zones résidentielles.

La figure 10 montre les résultats que donnent les mêmes méthodes que précédemment, mais en gardant à chaque fois 20 bandes. Dans ce cas, on constate que les deux dernières méthodes produisent un grand nombre de fausses alarmes concernant la zone de grands batiments, dont beaucoup sont détectés dans les zones résidentielles. D'autre part, comme on pouvait s'y attendre, le résultat de la méthode utilisant l'ACP à 20 composantes diffère très peu de celui à 10 composantes, puisque les bandes ajoutées sont de moins en moins informatives. La première méthode produit beaucoup moins de ces fausses alarmes, mais au prix d'une moins bonne homogénéité des zones de bâti industriel.

6.3 Temps de calcul

Le temps de calcul de la matrice de projection pour les différentes méthodes est très variable. Tout d'abord, notons que pour l'ACP, ce temps dépend du nombre de bandes de départ et d'arrivée, mais aussi du nombre de pixels de l'image, alors qu'en ce qui concerne la poursuite de projection, ce temps dépend du nombre de bandes d'arrivée et de départ, et du nombre de pixels d'apprentissage.

Dans le cas de la scène AVIRIS traitée au paragraphe précédent, de dimension spatiale 614×512, et ayant 224 bandes spectrales, les temps de calcul des matrices de projection correspondant aux différentes méthodes sont données dans le tableau de la figure 11.

Pour le cas de l'ACP, la longueur du temps de calcul s'explique aussi par le fait que l'image a des dimensions trop grandes pour être chargée en mémoire directement, et qu'il faut donc faire des lectures fréquentes sur le disque. La différence entre les deux méthodes de poursuite de projection s'explique par le fait que l'indice de projection est plus complexe dans le cas contextuel, puisqu'il s'agit d'une somme (ou d'un minimum) de l'indice sur les pixels d'apprentissage. D'autre part, le nombre d'itérations sur l'ensemble des bandes ainsi que le critère d'arrêt influencent les temps de calcul des algorithmes de poursuite de projection.

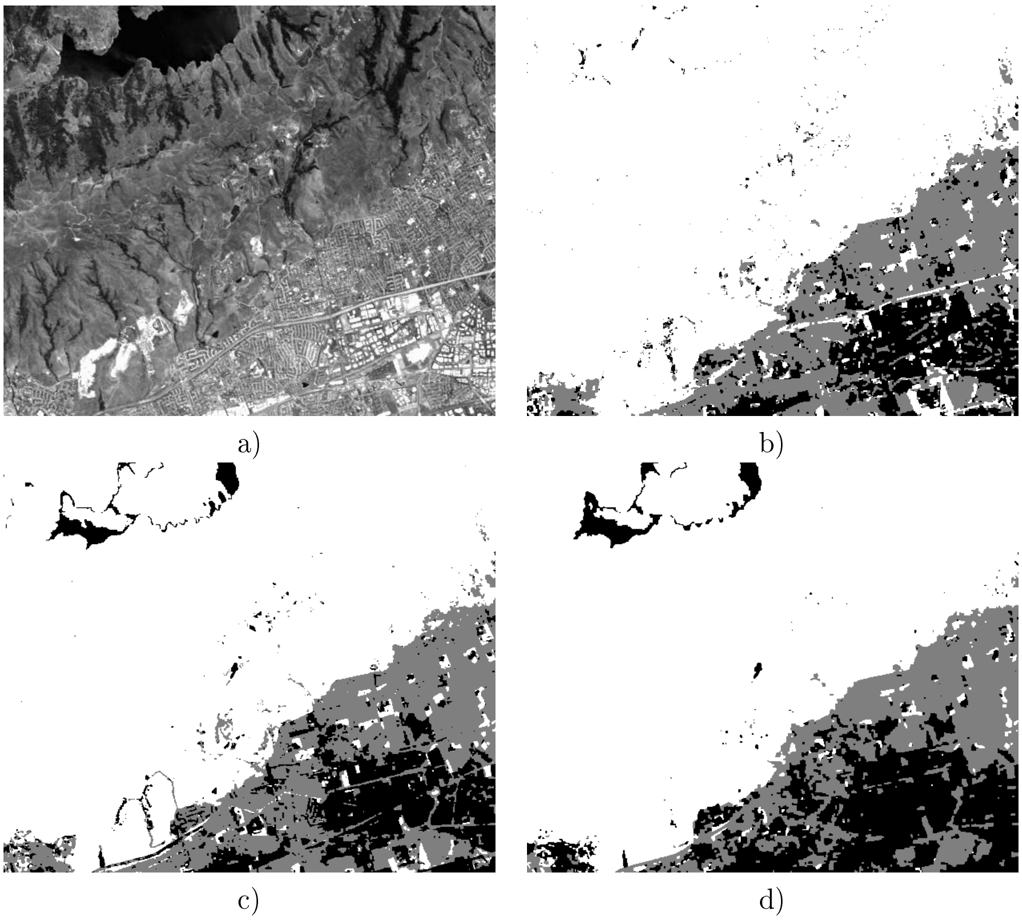


FIG. 9 – *a) Une bande de l'image initiale, b) résultat de classification sur 10 bandes extraites par poursuite de projection sur le modèle markovien gaussien, c) idem sur modèle gaussien pixélique, d) résultat de classification sur les 10 premières composantes principales, par maximum de vraisemblance sur modèle gaussien pixélique.*

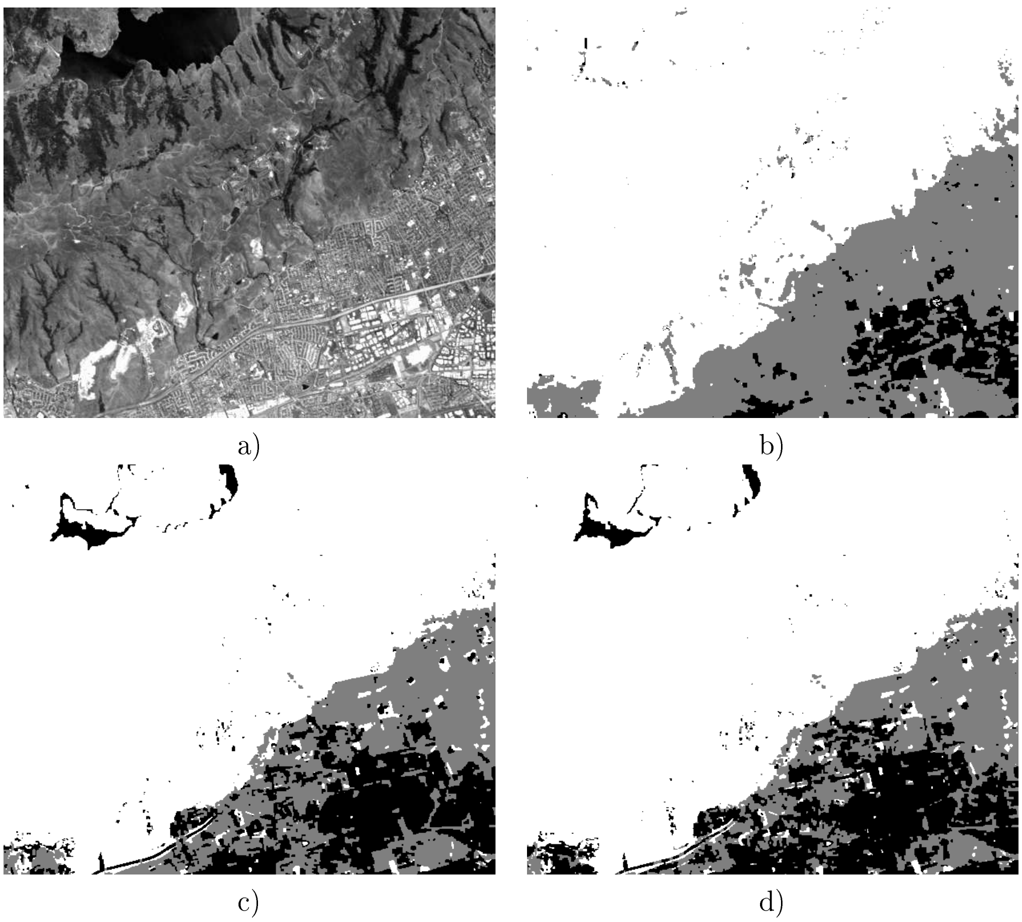


FIG. 10 – *a) Une bande de l'image initiale, b) résultat de classification sur 20 bandes extraites par poursuite de projection sur le modèle markovien gaussien, c) idem sur modèle gaussien pixélique, d) résultat de classification sur les 20 premières composantes principales, par maximum de vraisemblance sur un modèle gaussien pixélique.*

Type de réduction d'espace	Temps de calcul de la projection
ACP	1h25mn
PP-G	30mn
PP-MG	1h40mn

FIG. 11 – *Temps de calcul de la matrice de projection pour les 3 méthodes dans le cas de la scène AVIRIS de Moffett Field (sur une Sun Ultra10)*

7 Conclusion

Ce rapport présente une méthode de classification supervisée pour l'imagerie hyperspectrale, utilisant l'information de texture éventuellement présente dans les images.

Cette méthode est fondée sur une analyse de la texture à travers une modélisation globale par champ de Markov gaussien multivarié et sur l'utilisation d'un algorithme de poursuite de projection pour réduire la dimension de l'espace dans lequel est réalisée l'observation.

On a, dans un premier temps, présenté les raisons du passage de l'espace original à un espace de dimension réduite pour éviter les problèmes dus au phénomène de Hughes.

Dans une deuxième partie, on a présenté les méthodes les plus utilisées pour effectuer cette réduction de dimension, ainsi qu'un algorithme moins connu, la poursuite de projection, qui a la propriété d'être beaucoup moins sensible aux problèmes de dimensionalité, puisque les calculs y sont effectués dans le sous-espace projeté. On propose une variante de l'algorithme approché proposé par Jimenez et Landgrebe dans [7], qui consiste, étant donné un indice de projection caractérisant la discrimination entre les classes, à découper le spectre en groupes de bandes ayant des contributions homogènes à l'indice de projection, et d'extraire de chacun de ces groupes une composante, de telle sorte que l'indice de projection correspondant à l'espace généré par l'ensemble des composantes extraites soit maximal.

Enfin, la troisième partie a été consacrée à la modélisation de texture par champ de Markov gaussien, qui permet de prendre en compte la texture des images, et qui est assez souple puisque dans le cadre d'une image non-texturée, donc sans dépendance spatiale inter-pixel, on retrouve le cas gaussien pixélique. On propose une méthode d'estimation des paramètres de ce modèle dans le cadre d'une simplification due aux caractéristiques particulières des images hyperspectrales.

Cette modélisation est utilisée pour l'adaptation de l'algorithme de poursuite de projection, dans laquelle l'indice de projection à maximiser est lié à la distance de Bhattacharyya entre les classes, elle-même dépendante de la modélisation de celles-ci.

Les résultats présentés dans la dernière partie montrent plusieurs points notables :

- Premièrement, la distinction entre classes urbaines et rurales semble être facilement réalisée, quelle que soit la méthode utilisée.
- Deuxièmement, on voit que dans le cas d'images dont les classes sont bien séparées radiométriquement, et même en présence de texture, les résultats sont équivalents quand on utilise des méthodes classiques comme l'ACP, ou des méthodes moins classiques comme la poursuite de projection utilisant un modèle pixélique, ou utilisant le modèle contextuel proposé.
- Troisièmement, dans le cas de données réelles, on constate que les résultats paraissent meilleurs si l'on utilise le modèle contextuel, même si ceci peut difficilement être validé, en l'absence d'une réelle vérité de terrain faite par un expert. Toutefois, il semble que le modèle proposé soit plus capable de saisir toute la variabilité des tissus urbains, si l'on se fie aux cartes disponibles de la région.

En ce qui concerne les travaux futurs, il est possible d'adapter l'algorithme au cas non-supervisé. La comparaison avec d'autres modèles de texture est également envisagée.

Remerciements

Le premier auteur souhaite remercier la région Provence-Alpes-Côte d'Azur et Alcatel Space Industries pour le soutien financier apporté à sa thèse. Les auteurs remercient le JPL pour la mise à disposition de l'image AVIRIS de Moffett Field en Californie, et le Dr. Robin Morris du Centre de Recherche Ames de la Nasa pour les cartes fournies.

Références

[1] Haralick (R. M.). – Statistical and structural approaches to texture. *Proc of the IEEE*, vol. 67, n5, 1979, pp. 786–804.

[2] Jain (A. K.) et Farrokhnia (F.). – Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, vol. 24, n12, 1991, pp. 1167–1186.

[3] Schweizer (S. M.) et Moura (J.). – Hyperspectral imagery : Clutter adaptation in anomaly detection. *IEEE, Trans. on Information Theory*, vol. 46, n5, 2000, pp. 1855–1871.

[4] Harsanyi (J.C.) et Chang (C-I). – Hyperspectral image classification and dimensionality reduction : an orthogonal susbspace projection approach. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 32, n4, 1994, pp. 779–785.

[5] Thai (B.) et Healey (G.). – Using a linear subspace approach for invariant subpixel material identification in airborne hyperspectral imagery. *In : CVPR 99*, pp. 567–572.

[6] Hughes (G.F.). – On the mean accuracy of statistical pattern recognizers. *IEEE Trans. on Information Theory*, vol. 14, n1, 1968, pp. 55–63.

[7] Jimenez (L. O.) et Landgrebe (D. A.). – Hyperspectral data analysis and supervised feature reduction via projection pursuit. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, n6, 1999, pp. 2653–2667.

[8] Hall (P.) et Li (K.). – On almost linearity of low dimensionality projections from high dimensional data. *Ann. Stat.*, vol. 21, n2, 1993, pp. 867–889.

[9] Dunteman (G. H.). – *Principal Components Analysis.* – Sage Inc., 1989.

[10] Jia (X.) et Richards (J. A.). – Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 37, n1, 1999, pp. 538–542.

[11] Green (A.A.), Berman (M.), Switzer (P.) et Craig (M.D.). – A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 26, n 1, 1988, pp. 65–74.

[12] Fukunaga (K.). – *Introduction to statistical pattern recognition.* – Academic Press, 1990, 2eme édition.

[13] Huber (P. J.). – Projection pursuit. *The annals of Statistics*, vol. 13, n2, 1985, pp. 435–475.

[14] Lee (C.) et Choi (E.). – Bayes error evaluation of the gaussian ml classifier. *IEEE, Trans. on Geoscience and Remote Sensing*, vol. 38, n3, 2000, pp. 1471–1475.

[15] Descombes (X.). – *Champs de markoviens en analyse d'images.* – Thèse de Doctorat, ENST Paris, 1993.

[16] Winkler (G.). – *Image analysis, random fields and dynamic Monte Carlo methods : a mathematical introduction.* – Springer, 1995.

[17] Besag (J.). – Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Stat. Soc. B*, vol. 36, 1974, pp. 192–236.

[18] Derin (H.) et Elliott (H.). – Modeling and segmentation of noisy and textured images using Gibbs random fields. *IEEE. Trans. on Pattern Anal. and Mach. Intell.*, vol. 9, 1987, pp. 721–741.

[19] Descombes (X.), Sigelle (M.) et Préteux (F.). – Estimating Gaussian Markov random field parameters in a nonstationary framework : application to remote sensing imaging. *IEEE Trans. on Image Processing*, vol. 8, n 4, 1999, pp. 490–503.

[20] Lorette (A.), Descombes (X.) et Zerubia (J.). – Urban areas extraction based on texture analysis through a Markovian modelling. *Int. Journal of Computer Vision*, vol. 36, n 3, 2000, pp. 219–234.

[21] Panjwani (D. K.) et Healey (G.). – Markov random field models for unsupervised segmentation of textured color images. *IEEE, Trans. on Pattern Anal. Machine Intell.*, vol. 17, n 10, 1995, pp. 939–954.

[22] Hazel (G. G.). – Multivariate Gaussian MRF for multispectral scene segmentation and anomaly detection. *IEEE Trans. on Geoscience and Remote Sensing*, vol. 38, n 3, 2000, pp. 1199–1211.

[23] Jain (A. K.). – *Fundamentals of image processing*, chap. 2 : Two-dimensional systems and mathematical preliminaries. – Prentice Hall, 1989.

[24] Bennett (J.) et Khotanzad (A.). – Maximum likelihood estimation methods for multispectral random field image models. *IEEE, Trans. on Pattern Anal. and Machine Intell.*, vol. 21, n 6, 1999, pp. 537–543.

[25] Wu (F.Y). – The Potts model. *Reviews of modern physics*, vol. 54, n 1, 1982, pp. 235–267.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - B.P. 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - B.P. 101 - 54602 Villers lès Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot St Martin (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - B.P. 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, B.P. 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399