ANALYSE EN COMPOSANTES INDÉPENDANTES D'IMAGES MULTIBANDES : FAISABILITÉ ET PERSPECTIVES

Salim CHITROUB

Laboratoire de traitement du signal et d'images, Faculté d'électronique et d'informatique, Université des Sciences et Technologie Houari Boumediene, B.P. 32, El-Alia, Bab-Ezzouar, 16111 Alger, Algérie ;
Télécopieur : (213) (21) 24-76-07 et (213) (21) 24-71-87 ; courriel : s_chitroub@hotmail.com et schitroub@usthb.dz

Soumis le 8 mars 2007 ; révisé le 20 juin 2007 ; accepté le 1er août 2007 - © *Revue Télédétection*, 2007, vol. 7, n° 1-2-3-4, p. 541-555

Résumé

L'analyse en composantes indépendantes (ACI) est une méthode statistique d'analyse de données récemment développée pour traiter des observations vectorielles (multivariées) afin d'en extraire des composantes linéaires aussi indépendantes que possible. Cette technique émergente peut être effectuée tant par une approche algébrique que neuronale. Dans cet article, nous examinons la faisabilité et les perspectives de cette méthode pour l'analyse d'images multibandes de télédétection.

*Dans la communauté du traitement du signal, plusieurs algorithmes d'ACI ont été proposés. L'apprentissage d'ensemble est une méthode bayésienne récemment développée pour l'apprentissage des réseaux de neurones. Nous proposons donc d'effectuer l'ACI par l'apprentissage d'ensemble en vue d'obtenir une représentation optimale des images multibandes. Le modèle appliqué suppose que les pixels à séparer sont donnés sous forme de mélanges linéaires. Les matrices des mélanges linéaires sont supposées être inconnues. La méthode proposée est évaluée et testée sur une image multibande de 250 x 350 pixels fournie par le système d'acquisition d'images optiques ATM de six bandes spectrales. L'image représente une zone agricole de cinq classes, c'est-à-dire : carottes, pommes de terre, sol nu, chaume et betteraves à sucre. Afin d'effectuer une étude comparative, la méthode FastICA (*Fast Independent Component Analysis*), récemment développée (Hyvärinen, 1999) et basée sur le principe de l'algorithme d'apprentissage itératif de type point fixe, est appliquée sur la même image multibande. L'étude a montré que l'ACI par l'apprentissage d'ensemble donne des résultats acceptables par rapport aux résultats de la FastICA et sépare mieux les pixels tout en préservant leurs caractéristiques et les formes de l'image.*

Mots-clés : télédétection, image multibande, analyse en composantes principales, analyse en composantes indépendantes, séparation aveugle de sources, statistiques d'ordre supérieur, apprentissage d'ensemble, théorie de Bayes.

INDEPENDENT COMPONENT ANALYSIS OF MULTISPECTRAL IMAGES : FEASIBILITY AND PERSPECTIVES

Abstract

Independent Component Analysis (ICA) is a statistical method for data analysis recently developed to process multivariate data in order to extract the linear components that are as statistically independent as possible. This emergent technique can be carried out either by algebraic approaches or by neuronal approaches. We propose in this paper to study the feasibility and the potential of this method for remote sensing multispectral image analysis. In the signal processing community, several ICA algorithms were proposed. The Ensemble Learning (EL) is a Bayesian method recently developed for neural network training. Here, we propose to carry out an ICA based on EL for remote sensing multispectral image analysis. The applied model is as follows. A multispectral image is supposed a collection of multiple pixels of size m *(number of spectral components)* x(r) *indexed by the position* r. *We suppose that the pixels to be separated are given in the form of linear mixtures. The matrices of the linear mixtures are supposed to be unknown. Such a separation is known as blind. The suggested method is evaluated and tested on a multispectral image of 250 x 350 pixels provided by the ATM six spectral band optical imaging system. The image represents an agricultural scene containing five classes. In order to carry out a comparative study, the Fast Independent Component Analysis (FastICA)*

method, recently developed in the literature and based on the principle of the iterative fixed point algorithm, is applied on the same multispectral image. The study showed that the EL-based ICA gives acceptable results compared to the results of FastICA and separates the pixels more accurately while preserving the characteristics and forms of the imaged scene.

Keywords : remote sensing, multispectral image, principal component analysis, independent component analysis, blind source separation, higher order statistics, ensemble learning, Bays theory.

1. INTRODUCTION

La séparation aveugle de sources, dite BSS (Blind Source Separation), est un domaine en plein développement dans la communauté du traitement du signal et plusieurs méthodes, statistiques et neuronales, ont été proposées pour trouver une solution analytique performante à ce problème. Supposons que l'on dispose de plusieurs capteurs recevant des signaux de sources distinctes. Chaque capteur reçoit un mélange linéaire des signaux sources : on appelle ce mélange signal mesuré. Cependant, ce mélange linéaire n'est pas connu. Mais, avec les méthodes de séparation de sources, on tente de reconstituer les signaux sources. On dit que la séparation est aveugle en raison du fait que l'on ne dispose d'aucune information propre aux signaux sources ni de la forme de la combinaison linéaire des signaux reçus.

Ce problème a vu le jour vers la fin des années 1980. Les premiers auteurs à l'avoir exposé sont Herault et Jutten (Jutten, 1987) dans leurs travaux neurobiologiques qui consistaient à étudier les réponses musculaires émises en fonction de différentes sortes d'excitation. Ensuite, plusieurs travaux ont vu le jour qui ont abouti à l'ACI. L'ACI est une technique statistique dont l'objectif est de décomposer un signal aléatoire multivariable X (le signal mesuré) en une combinaison linéaire de signaux indépendants (les signaux sources), pour faire ressortir des signaux aussi indépendants que possible à partir des signaux mesurés.

Soit, alors X = A.S, où S est un signal à composantes indépendantes et A une matrice à coefficients réels. Grâce au théorème de Darmois (Darmois, 1953), Comon (Comon, 1994) a démontré que l'ACI peut être une solution au problème de la BSS, sous certaines conditions. Dans le cas où il y aurait autant de sources S que d'observations X, le problème de l'ACI peut être reformulé de la façon suivante : trouver une matrice carrée W (qui correspond à A^{-1}) telle que le vecteur obtenu par action de W sur les observations X aient des composantes les plus indépendantes possible (Cardoso, 1999).

Plusieurs travaux sur l'ACI ont été menés par divers chercheurs. D'une manière générale, ces travaux peuvent se classer en deux catégories de méthodes : les méthodes basées sur les concepts algébriques en faisant appel aux statistiques d'ordre supérieur (Comon, 1994 ; Cardoso, 1999 ; Achard, 2003) et les méthodes neuronales dans lesquelles l'ACI est reformulée sous forme de règles d'apprentissage non dirigées (Hyvärinen *et al*., 2001). La convergence du réseau de neurones correspondant permet d'obtenir l'inverse de la matrice de mélange linéaire de départ (Hyvärinen, 1999 ; Lee *et al*., 2000). Dans cet article, nous avons adopté l'approche neuronale pour l'ACI en utilisant la méthode bayésienne dans le processus d'apprentissage du réseau de neurones. Ce type d'apprentissage est appelé apprentissage d'ensemble (AE), ou EL (Ensemble Learning) en anglais, récemment développé (Mackay, 1995 ; Neal, 1996 ; Lappalainen and Miskin, 2000 ; Lappalainen and Honkela, 2000).

Des chercheurs ont déjà abordé la théorie et l'application de l'ACI pour l'analyse et le traitement de données de télédétection (Bayliss *et al*., 1998 ; Zhang and Chen, 2002 ; Lennon, 2002 ; Botchko *et al*., 2003 ; Nascimento and Bioucas Dias, 2005 ; Wang and Chang, 2006a, 2006b). Au cours des dernières années, l'intérêt qui a été accordé à l'ACI pour résoudre les problèmes des images radar, en particulier si le chatoiement est considéré comme étant un bruit multiplicatif de statistiques non gaussiennes (Chitroub *et al*., 2000, 2002b ; Chitroub, 2003, 2006). En revanche, l'application de l'ACI pour l'analyse des images multibandes optiques est timidement abordée en raison de l'absence de cohérence entre la théorie et les hypothèses de départ de l'ACI ainsi que le modèle mathématique des images multibandes optiques.

Dans le reste de l'article, on présente d'abord le problème de la BSS et comment l'ACI peut être considérée comme solution au problème de la BSS. On présente ensuite la faisabilité de l'ACI en télédétection optique et ses perspectives d'application ainsi que l'étude détaillée de l'ACI par l'EL en vue d'obtenir une représentation optimale des images multibandes, avec les résultats obtenus commentés. Enfin, une étude comparative est effectuée avec la méthode de la FastICA récemment développée.

2. SÉPARATION AVEUGLE DE SOURCES

Soit *S* un vecteur aléatoire tel que $S = \left(s_1, s_2, \ldots s_n\right)^T$ où les s_i avec $\left(i = 1, \ldots, n\right)$ sont des signaux aléatoires statistiquement indépendants, appelés sources, ayant des distributions de probabilité inconnues *a priori*. Ce sont en fait les signaux indépendants émis par les différentes sources. Dans notre cas, les sources sont les images spectrales constituant l'image multibande. Soient *m* capteurs, apportant *m* mesures sur les combinaisons linéaires des *n* signaux sources originaux et constituant le vecteur des signaux mesurés $X = \left(x_1, x_2, \ldots x_m\right)^T$ tel que :

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

Le but est donc de retrouver la matrice de séparation W tel que :

$$S = W.X \quad (2)$$

La séparation est dite aveugle puisqu'on n'utilise que le vecteur des signaux mesurés à la sortie des capteurs pour obtenir les signaux sources originaux émis sans aucune connaissance *a priori* sur ces derniers. Sur la figure 1, on présente un schéma simplifié d'un système de séparation aveugle de sources, où X est le vecteur des signaux mesurés sous forme d'un mélange linéaire (c'est la seule information disponible), S étant le vecteur des signaux sources originaux, $\hat{S}$ étant le vecteur des signaux sources estimés à l'aide d'un algorithme de séparation aveugle de sources et A est la matrice de mélange inconnue tel que $W \approx A^{-1}$, et ce à un facteur près que nous tentons de trouver. En combinant les équations (1) et (2), nous obtenons la matrice du système global de séparation aveugle de sources, d'ordre mxn et notée par C, tel que :

$$C = A.W \quad (3)$$

Sur la figure 2, on représente le schéma global du système de séparation aveugle de sources (Cikochi and Amari, 2002). Nous remarquons que, dans un cas idéal où $\hat{S}$ serait égal à S, la matrice C trouvée serait égale à la matrice identité.

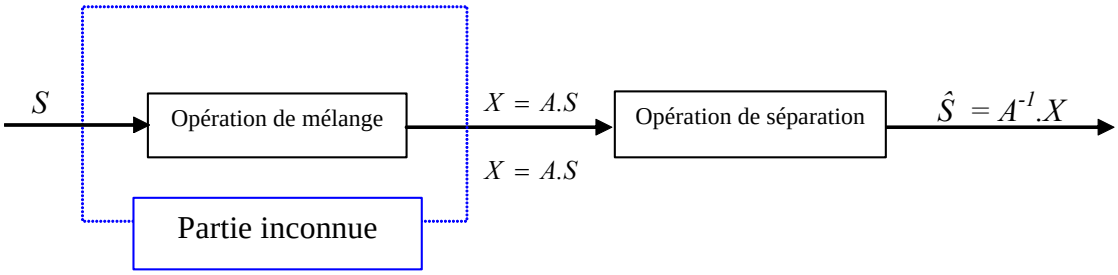


Figure 1 : Schéma simplifié d'un système de séparation de sources. *Simplified diagram of source separation system.*

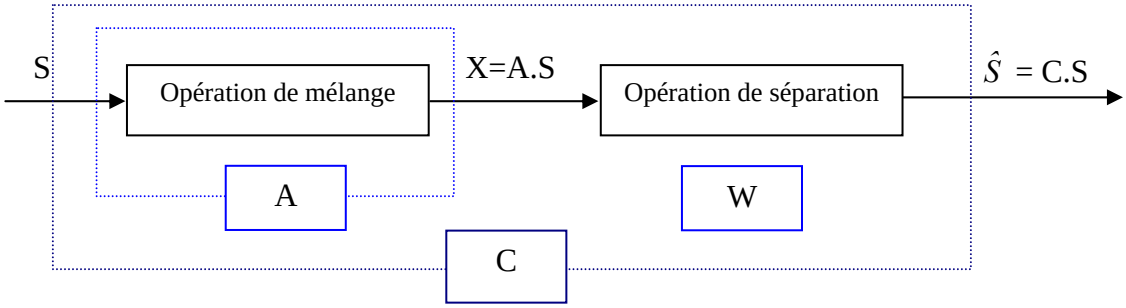


Figure 2 : Schéma global d'un système de séparation de sources. *Global diagram of source separation system.*

3. ANALYSE EN COMPOSANTES INDÉPENDANTES

L'ACI est une méthode statistique qui traite des observations vectorielles (multivariées) afin d'en extraire des composantes linéaires aussi indépendantes que possible. Cette idée simple s'est révélée très fructueuse pour le traitement des signaux et des images multi-capteurs dans de nombreux domaines tels que : les interférences des ondes radio, les réseaux d'antennes pour les télécommunications, la prise de son, les signaux biomédicaux multi-électrodes, l'imagerie biomédicale, les images de télédétection (optiques et radar) (Ristaniemi and Joutsensalo, 1999 ; Zhang and Chen, 2002 ; Botchko *et al*., 2003 ; Nascimento and Bioucas Dias, 2005 ; Loghmari *et al*., 2006 ; Wang and Chang, 2006a, b ; Chitroub *et al*., 2000, 2002b ; Chitroub, 2004, 2006). De manière plus générale, c'est utilisé dans tous les cas où un système de plusieurs capteurs, fournissant des signaux (ou images) cohérents, est à l'écoute d'un ensemble discret de sources que l'on cherche à extraire des observations et que l'on peut, pour des raisons physiques, supposer mutuellement indépendantes statistiquement.

Cependant, l'idée de rechercher des composantes indépendantes était trop générale pour rester confinée dans son domaine d'origine : le traitement des signaux et images. L'ACI est en réalité une extension d'une méthode d'analyse de données très connue à savoir l'analyse en composantes principales (ACP). En fait, si l'ACP cherche à extraire des variables décorrélées, en se limitant à imposer la contrainte d'indépendance aux statistiques d'ordre deux (matrice de covariance), l'ACI en revanche, cherche l'indépendance des statistiques d'ordre supérieur à deux des variables mesurées. L'ACI est aussi connectée, en tant que méthode d'exploration de données, aux idées de la méthode dite projection-poursuite (PP) (Hyvärinen *et al*., 2001). En utilisant la méthode PP, on cherche des directions intéressantes dans un nuage de données multidimensionnelles, en arguant du fait que les structures intéressantes sont celles qui font apparaître le plus l'aspect non gaussien des données projetées.

Il existe un lien très significatif entre dépendance, corrélation et l'aspect non gaussien d'une variable aléatoire dans le cadre du modèle linéaire de l'ACI. En fait, la corrélation ne dépend que de la matrice de covariance. Elle représente, en quelque sorte, la partie gaussienne de l'information mutuelle qui existe entre la vraie distribution de la variable aléatoire et sa distribution estimée. L'aspect non gaussien d'une variable aléatoire est la divergence qui existe entre sa distribution et sa meilleure approximation gaussienne. Ainsi, l'ACI, qui cherche à expliquer une variable aléatoire multidimensionnelle *X*, en termes de composantes linéairement et statistiquement indépendantes, peut être reformulée comme étant la minimisation de la corrélation entre les composantes de X tout en maximisant l'aspect non gaussien de chaque composante (et non l'aspect non gaussien marginal). Dans les sous-sections suivantes, nous donnons les mesures de l'indépendance statistique qui peuvent être utilisées dans la mise en œuvre de l'ACI. Ces mesures sont utilisées dans l'élaboration des critères d'optimisation des algorithmes de l'ACI. Ces critères sont appelés fonctions de contrastes (Comon, 1994).

3.1. Divergence de Kullback-Leibler

La divergence de Kullback-Leibler permet de mesurer l'écart qui existe entre la densité conjointe et le produit des densités marginales. Cet écart est la mesure d'indépendance statistique (Kullback and Leibler, 1951). Notons que les composantes de la variable aléatoire $X = \left(x_1, x_2, \ldots x_m\right)^T$ sont dites mutuellement indépendantes statistiquement si et seulement si :

$$P(X) = \prod_{i=1}^{m} P(x_i) \text{ pour chaque paire } i, j = 1 \ldots m \qquad (4)$$

où $P(X)$ est la distribution de X et $\prod_{i=1}^{m} P(x_i)$ est le produit de ses distributions marginales.

Soit *f(x)* et *g(x)* deux densités de probabilités des deux composantes de la variable aléatoire X. Ainsi, la divergence de Kullback-Leibler est définie par :

$$D_{f,g} = \int_{-\infty}^{+\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \qquad (5)$$

Il est clair que la divergence de Kullback-Leibler ne s'annule que lorsque les deux composantes de X sont statistiquement indépendantes.

3.2. Information mutuelle

L'information mutuelle peut être considérée comme la distance de Kullback-Leibler entre les densités conjointes et les densités marginales (Shannon, 1948 ; Kullback and Leibler, 1951). Ainsi, l'information mutuelle de la variable aléatoire X, notée *I(X)*, peut être exprimée comme étant la divergence de Kullback-Leibler entre sa distribution *P(X)* et le produit de ses distributions marginales $P(x_i)$ tel que :

$$I(X) = D\left(P(X) \middle| \prod_{i=1}^{m} P(x_i)\right) \qquad (6)$$

En utilisant le concept de l'entropie différentielle (Cover and Thomas, 1991), on peut reformuler l'information mutuelle *I(X)* comme suit :

$$I(X) = I(x_1, \ldots, x_m) = \sum_{i=1}^{m} H(x_i) - H(X) \qquad (7)$$

où les $H(x_i)$ sont les entropies marginales des composantes x_i et $H(X)$ l'entropie conjointe.

4. ACI EN TÉLÉDÉTECTION : FAISABILITÉ ET PERSPECTIVES

Bien que l'application de l'ACI à l'imagerie radar soit mathématiquement justifiée, à travers l'hypothèse de bruit multiplicatif non gaussienne pour le chatoiement et les modèles statistiques caractérisés par les statistiques d'ordre supérieur pour les images polarimétriques mono- et multi-vue (Chitroub *et al*., 2002a), en imagerie optique multibande, cependant, les choses ne sont pas encore claires et ceci pour des raisons que nous évoquons dans les sous-sections suivantes.

4.1. Problématiques des images multibandes

Les algorithmes de classification multibandes développés donnent des résultats satisfaisants si la dimension des données ne dépasse pas une dizaine de bandes spectrales. Si la dimension est au-delà de dix, les résultats se dégradent rapidement à cause du phénomène de Hughes (Hughes, 1968), qui se manifeste par une perte de précision dans l'estimation des paramètres des densités de probabilités multidimensionnelles des classes. Ce problème, bien connu en télédétection, peut être abordé en effectuant au préalable une étape de réduction de l'espace de représentation des données multibandes (Landgrebe, 1999). En effet, la plupart des objets terrestres émettent ou reflètent l'énergie sur une grande partie du spectre (Bonn et Rochon, 1992). Ceci se traduit habituellement par l'existence de corrélations significatives entre les différentes bandes spectrales et donc une redondance d'information. Il est donc souhaitable d'éliminer les corrélations afin d'améliorer les résultats de la classification.

De plus, la projection géométrique d'un élément simple d'un capteur (le pixel) sur la surface au sol, appelée champ visuel instantané, dit IFOV (*Instantaneous Field Of View*), est constituée d'un mélange des signatures spectrales en raison de l'hétérogénéité de la plupart des surfaces au sol. D'après Campbell (Campbell, 1996), cette situation surgit quand la cellule de résolution, c'est-à-dire la surface au sol définie par l'IFOV, est supérieure à la résolution du capteur. Ainsi, le pixel mixte se produit en raison d'une résolution spectrale élevée ou d'une basse résolution spatiale. Les substances matérielles qui constituent le contenu du pixel mixte sont appelées éléments composants ou membres extrêmes (en anglais : *Endmembers*). Un membre extrême représente donc une classe et, par conséquent, extraire les membres extrêmes est équivalent à trouver les différentes classes. Les méthodes classiques d'analyse et de classification des images optiques, qui considèrent que le pixel est pur, ne peuvent pas s'appliquer pour analyser les images multibandes de dimension élevée et pour classifier les pixels mixtes.

4.2. Application de l'ACI : hypothèses et cohérence

Les premiers travaux sur l'application de l'ACI en télédétection optique ont été orientés vers la recherche d'une solution au problème des pixels mixtes (Bayliss *et al*., 1998 ; Parra *et al*., 2000 ; Zhang and Chen, 2002 ; Botchko *et al*., 2003 ; Nascimento and Bioucas Dias, 2005). Ces travaux ont proposé d'extraire les membres extrêmes et d'estimer leurs signatures spectrales ainsi que leurs fractions (appelé fractions d'abondance), en utilisant seulement les données observées (pixels mixtes), d'où l'intérêt d'utiliser l'ACI. Les deux hypothèses de départ dans l'application de l'ACI dans ces travaux sont : 1) le contenu de chaque pixel mixte est considéré comme étant un mélange linéaire des réponses spectrales (sources) des membres extrêmes, dont les coefficients sont les fractions d'abondance correspondantes ; 2) les sources sont supposées être statistiquement indépendantes.

La première hypothèse est vérifiée puisque la dispersion multiple entre les membres extrêmes distincts est négligeable et la surface du pixel est partagée entre les abondances partielles propres aux membres extrêmes. Cependant, la deuxième hypothèse est inadmissible due aux contraintes physiques du procédé d'acquisition. En fait, la somme des fractions d'abondance associées à chaque pixel mixte est constante et, par conséquent, les réponses spectrales (sources) des membres extrêmes, qui constituent le mélange linéaire, ne peuvent pas être statistiquement indépendantes (Settle and Drake, 1993). En d'autres termes, ces travaux sont principalement concentrés sur la séparation de sources (réponses spectrales des membres extrêmes). La quantification d'abondance des sources est timidement abordée et exige plus de développement. Le problème important de la détermination du nombre de sources à séparer était aussi absent dans ces travaux. Ce problème demeure non résolu et il est d'intérêt dans la communauté de traitement du signal.

4.3. ACI et réduction de la dimension

L'ACP n'est pas appropriée pour l'analyse des images multibandes de dimension élevée. La transformation de fraction maximale de bruit, dite MNF (*Maximum Noise Fraction Transform*) (Green *et al*., 1988), et l'ACP à bruit ajusté, dite NAPC *(Noise-Adjusted Principal Components*) (Lee *et al*., 1999), souffrent également des mêmes inconvénients. En fait, le problème qui se pose avec l'ACP, la MNF et la NAPC est que ces méthodes sont basées sur la décomposition en valeurs propres des matrices de covariance des images et du bruit additif (Chitroub, 2003). Cependant, les images multibandes de dimension élevée contiennent des informations spectrales riches sur les différents objets terrestres. Cette richesse dans l'information spectrale ne peut pas être entièrement caractérisée par les statistiques de second ordre. Ceci est dû au fait que les échantillons extraits des différentes cibles sont relativement insuffisants pour constituer des statistiques fiables. Ainsi, ces cibles ne peuvent pas être caractérisées par l'ACP, la MNF ou la NAPC.

Un autre problème posé avec ces méthodes est la détermination du nombre de composantes images à retenir. Le critère utilisé pour résoudre ce problème est de sélectionner les composantes dont les valeurs propres correspondantes représentent un certain pourcentage d'énergie requis pour la reconstitution des images originales. Mais, tel qu'il a été

démontré par Chang et Du (2004) et Ramakishna *et al.* (2005), ce critère n'était pas une mesure efficace pour les images multibandes de dimension élevée.

L'utilisation de l'ACI pour réduire la dimension des images multibandes a suscité peu d'attention dans le passé puisque que l'ACI n'a pas été développée à l'origine pour résoudre ce problème. Récemment, l'ACI a été utilisée pour obtenir une meilleure représentation des images multibandes en vue d'obtenir une classification plus précise de la couverture terrestre (Loghmari *et al*., 2006). Notons aussi que cette issue a été utilisée par Wang et Chang (2006a) où un nouveau concept présenté dans Ramakishna *et al*., 2005, appelé la dimension virtuelle ou DV (*Virtual Dimensionality*), a été utilisé pour estimer le nombre de composantes indépendantes requis pour une réduction de la dimension sans perte d'information des images originales. En fait, contrairement à l'ACP, où les composantes principales sont extraites en fonction de l'importance de la variance (valeurs propres), l'ACI ne donne pas des priorités aux composantes indépendantes à extraire. Plus précisément, les premières composantes indépendantes produites ne sont pas nécessairement plus importantes ou significatives que celles produites en dernier. Ce dilemme est engendré par le fait qu'il n'y a aucun critère établi pour extraire les composantes indépendantes selon une priorité donnée, comme c'est le cas de l'ACP, où les valeurs propres de la matrice de covariance sont utilisées pour donner la priorité aux premières composantes principales extraites. Puisque l'ACI utilise généralement des vecteurs aléatoires, en tant que vecteurs initiaux de projection pour initialiser l'algorithme d'extraction des composantes, l'ordre dans lequel les composantes sont extraites par la suite est donc complètement aléatoire.

4.4. ACI et extraction des membres extrêmes

Plusieurs algorithmes d'extraction des membres extrêmes, dits EEA (*Endmember Extraction Algorithm*), ont été développés par divers chercheurs. On peut ainsi citer par exemple, l'algorithme du N-trouveurs (*N-finder*), l'index de pureté de pixel ou PPI (*Pixel Purity Index*), l'analyse itérative d'erreur ou IEA (*Iterative Error Analysis*), l'algorithme d'extraction morphologique automatisée des membres extrêmes ou AMEE (*Automated Morphological Endmember Extraction*), la transformation de volume minimum, la géométrie convexe, l'analyse convexe de cône, l'analyse en composantes de sommet ou VCA (*Vertex Component Analysis*), etc. (Winter, 1999 ; Wang and Chang, 2006a ; Ramakishna *et al*., 2005 ; Chang and Plaza, 2006 ; Wang and Chang, 2006b). Toutes ces approches partagent un élément en commun qui est l'utilisation de la convexité pour déterminer les membres extrêmes. Cependant, elles souffrent également d'un inconvénient commun, qui est l'absence d'un guide directif pour déterminer le nombre exact de membres extrêmes à extraire. En effet, les algorithmes EEA identifient seulement les membres extrêmes qui sont supposés être purs et, par conséquent, ils ne donnent aucune information ni sur le nombre exact de ces membres ni sur l'estimation de leurs fractions des abondances correspondantes. Pour que l'algorithme EEA achève également la quantification d'abondance, l'algorithme de quantification d'abondance, tel que l'analyse linéaire de mélange au sens des moindres carrés ou LSMA (*Least Squares Linear Mixture Analysis*) (Chang *et al*., 2002 ; Ramakishna *et al*., 2005) est utilisée après l'extraction des membres extrêmes. On trouve aussi le modèle spatial de mélange linéaire ou LMS (*Linear Mixture Spatial*) (Loghmari *et al*., 2006), qui a été proposé pour estimer les membres extrêmes. L'idée est de relier les signatures des classes spectrales dans l'IFOV aux signatures des différents éléments composants du LMS, sachant que les variables aléatoires associées aux éléments composants sont supposées être statistiquement indépendantes. Cependant, l'inconvénient de cette approche réside dans la connaissance *a priori* des signatures spectrales des constituants du mélange qui est une tâche très délicate dans la pratique.

Contrairement à la méthode du LMS, l'ACI peut être utilisée pour récupérer les éléments composants à partir des pixels des images qui sont données sous forme d'un mélange linéaire inconnu, constitué des différentes réponses spectrales des objets terrestres dans l'IFOV. Ceci peut être réalisé en reconstituant l'indépendance statistique entres les membres extrêmes. Ainsi, l'absence de la connaissance *a priori* sur le mélange est compensée par cette forte, mais souvent physiquement plausible, hypothèse statistique. Plus précisément, si les signatures spectrales pures sont considérées comme des sources, l'extraction des membres extrêmes semble normale et justifiable comme application de l'ACI. En effet, puisque les membres extrêmes distincts représentent différentes classes, leur dépendance statistique est donc très faible. De plus, en raison de la nature des membres extrêmes en tant que signatures pures, la probabilité de l'occurrence des membres extrêmes est généralement faible et leur présence spatiale est également souvent peu fréquente. Tous ces profils semblent s'adapter à l'esprit de l'ACI (Hyvärinen *et al*., 2001) qui utilise l'information mutuelle pour mesurer l'indépendance statistique.

Cependant, comme dans le cas de la réduction de la dimension, l'extraction des membres extrêmes en utilisant l'ACI est confrontée au problème de l'identification du nombre des composantes à extraire. Ce problème est critique puisqu'on ne connaît pas au préalable les composantes indépendantes qui caractérisent les membres extrêmes. Dans ces conditions, toutes les composantes indépendantes sont extraites afin d'éviter une éventuelle perte d'un membre extrême. Malheureusement, dans le cas d'une dimension élevée, il y a généralement des centaines de bandes spectrales, dont chacune correspond à une image composante, et le choix des composantes indépendantes devient donc crucial. En fait, parmi ces centaines des composantes, seulement une dizaine peuvent contenir une information pertinente concernant les membres extrêmes.

Pour faire face à ce problème, deux algorithmes ont été développés par Wang et Chang (2006a). Le premier, HOS-ICPA (*Higher-Order Statistics-based Independent Component Prioritization Algorithm*), fait appel aux statistiques

d'ordre supérieur pour classer les composantes indépendantes à extraire. Le deuxième algorithme est une sorte d'orientation du processus d'apprentissage d'un algorithme quelconque de l'ACI. L'idée de base consiste à guider le début du processus d'extraction des composantes par un choix optimal de l'ensemble de vecteurs de projection initiaux afin de faire l'extraction des composantes selon le même ordre dans lequel la base initiale de projection a été générée. Avec l'utilisation d'un tel algorithme d'initialisation, l'aspect aléatoire dans l'extraction des composantes indépendantes peut être évité.

Mais le problème qui se pose avec ces deux algorithmes est l'estimation de la fraction d'abondance de chaque membre extrême extrait. En effet, afin de rendre l'ACI apte à quantifier l'abondance, elle doit satisfaire deux contraintes : la somme des valeurs des abondances doit être égale à un (ASC : *Abundance Sum-to-one Constraint*) et les abondances ne doivent pas être négatives (ANC : *Abundance Non-negativity Constraint*). Cependant, l'ACI utilise l'indépendance statistique comme critère de séparation de sources et, par conséquent, les résultats de l'ACI ne sont pas nécessairement positifs, d'autant plus que l'échelle est invariante. Ainsi, les deux contraintes (ASC et ANC) n'ont pas été respectées par l'ACI. Cette réalité est due au fait que, si l'ACI est appliquée en tenant compte des contraintes d'abondance, les membres extrêmes ne sont pas donc statistiquement indépendants. Ainsi, l'application de l'ACI pour la quantification d'abondance n'a jamais était atteinte.

Néanmoins, les algorithmes de l'ACI peuvent être forcés pour effectuer en même temps l'extraction des membres extrêmes et la quantification d'abondance sans utiliser un autre algorithme pour la quantification d'abondance. Ceci est possible si l'extraction des composantes indépendantes sélectionnées pour la réduction de la dimension est orientée vers les membres extrêmes. Ainsi, puisque chaque membre extrême représente une classe distincte, on peut donc imposer que chaque membre extrême soit extrait dans une composante indépendante à part (Wang and Chang, 2006b).

5. ACI PAR APPRENTISSAGE D'ENSEMBLE ET ANALYSE D'IMAGES MULTIBANDES OPTIQUES

Le problème de la BSS peut être reformulé comme étant un problème de séparation entre les distributions de mélange *a posteriori* des sources. La séparation des distributions sert, en fait, à les rendre statistiquement indépendantes. Ainsi, la BSS peut être résolue en utilisant l'ACI. Dans le cas de l'application présentée dans cet article, il est plus cohérent de raisonner en termes de séparation entre les distributions *a posteriori* qui caractérisent l'image multibande et d'effectuer la séparation des distributions dans le cadre bayésien de l'ACI. Ceci sera abordé à la section 5.3. Cependant, avant de détailler comment le problème de la BSS peut être résolu en utilisant l'ACI, nous allons expliquer que la séparation des distributions *a posteriori* d'un mélange consiste à les rendre statistiquement indépendantes. Ceci nous ramène à aborder l'approche bayésienne de l'ACI. Cette approche est basée sur l'AE, qui est une méthode bayésienne récemment développée pour l'apprentissage des réseaux de neurones (Mackay, 1995 ; Neal, 1996 ; Lappalainen and Miskin, 2000 ; Lappalainen and Honkela, 2000).

5.1. Apprentissage d'ensemble (AE)

Dans le cadre bayésien, l'information recherchée est contenue dans les distributions conditionnelles des observations. Le théorème de Bayes permet d'évaluer les distributions conditionnelles des informations inconnues, y_i, sachant l'ensemble des données observées, x_i. La distribution *a priori* $p(y)$ contient l'information concernant les variables inconnues avant l'observation. De son côté, la distribution *a posteriori* $p(y|x)$ contient l'information du système après l'observation. La vraisemblance $p(x|y)$ est la probabilité des données observées suivant un ensemble de valeurs des paramètres inconnus. Dans la pratique, la densité conjointe $p(y,x)$ et la distribution $p(x)$, sont difficiles à évaluer de même que la marginalisation de la distribution *a posteriori*. Par conséquent, il est nécessaire d'estimer la densité *a posteriori* par une forme plus simple afin de rendre possible la séparation entre les distributions de mélange.

L'estimateur du maximum *a posteriori* (MAP) mène à une sous-estimation ajustée de la précision. En fait, le MAP ne garantit pas une masse élevée de probabilité dans la crête de la distribution. Or, le but est de maintenir la masse de distribution *a posteriori*. L'estimateur de maximum de vraisemblance (MV) peut souffrir d'une sous-estimation des données si le modèle choisi n'est pas assez spécifique aux observations. Des modèles plus complexes peuvent donner une meilleure vraisemblance mais, si de nouvelles données sont rajoutées, l'estimateur ne pourra être généralisé par ces nouvelles données.

L'AE permet d'obtenir un modèle qui caractérise fidèlement la masse élevée de la distribution *a posteriori*. L'AE suppose que la distribution *a posteriori* réelle admet une décomposition en plusieurs distributions paramétriques. Les formes paramétriques des distributions sont utilisées dans la construction de la fonction de coût d'apprentissage à minimiser. Un choix convenable des formes paramétriques des facteurs de distribution permet d'optimiser le processus d'apprentissage et rend l'estimation des distributions séparées plus précise.

5. 2. ACI par l'AE

La méthode de l'AE est utilisée dans le processus d'apprentissage de l'ACI afin de séparer les distributions *a posteriori* du mélange mesuré. Elle permet de faire l'approximation paramétrique des distributions des observations *a posteriori* où l'information de Kullback-Leibler est utilisée pour la mesure de la différence entre la distribution *a posteriori* actuelle et son approximation. Le déroulement de l'algorithme se fait d'une manière itérative de sorte que la distribution *a posteriori* d'un cycle devient la distribution *a priori* dans le cycle suivant.

L'approximation de la distribution $p(y|x)$ par la distribution $q(y|x)$ tout en minimisant la mesure de Kullback-Leibler est telle que :

$$\begin{aligned} D\left(q(y|x)\|p(y|x)\right) &= \int_{-\infty}^{+\infty} q(y|x)\log\left(\frac{q(y|x)}{p(y|x)}\right)dy \\ &= \int_{-\infty}^{+\infty} q(y|x)\log\left(\frac{p(x)q(y|x)}{p(x,y)}\right)dy \\ &= \int_{-\infty}^{+\infty} q(y|x)\log\left(\frac{q(y|x)}{p(x,y)}\right)dy + \log(p(x)) \end{aligned} \tag{8}$$

La valeur de la mesure de Kullback-Leibler est positive ou nulle. Dans le cas où elle est nulle, les deux distributions $p(y|x)$ et $q(y|x)$ sont équivalentes. Cependant, dans la pratique, cette valeur est difficile à annuler, mais on cherche à la minimiser. Le terme $p(x)$ est invariant durant l'apprentissage : on pourra donc définir une fonction de contraste (ou fonction du coût) $C_y(x)$ dont le minimum global caractérise la distribution estimée optimale :

$$C_y(x) = D\left(q(y|x)\|p(y|x)\right) - \log(p(x)) = \int_{-\infty}^{+\infty} q(y|x)\log\left(\frac{q(y|x)}{p(x,y)}\right)dy \tag{9}$$

L'apprentissage d'ensemble est pratique, si les termes $p(y,x)$ et $q(y|x)$ de la fonction de coût $C_y(x)$ peuvent être factorisés. Si c'est le cas, les logarithmes dans la fonction deviennent des sommes de termes simples. En vertu de la définition des modèles, la vraisemblance et les probabilités *a priori* peuvent être des produits des distributions.

L'information de Kullback-Leibler est une mesure globale, sachant que la distribution estimée est une distribution globale. Par conséquent, la mesure sera sensible à la masse de probabilité dans la vraie distribution *a posteriori* plutôt qu'à la valeur absolue de la distribution elle-même. Dans l'hypothèse que l'ensemble possède une forme paramétrique fixe, la fonction de coût est alors minimisée sachant les paramètres des distributions.

5.3. ACI par l'AE pour l'analyse d'images multibandes optiques

À partir du raisonnement de l'ACI par l'AE, nous pouvons pallier le problème de cohérence entre le fondement théorique de l'ACI classique et son application pour l'analyse des images multibandes optiques. En fait, l'hypothèse de l'aspect non gaussien des modèles statistiques des sources requis par l'ACI est allégée avec l'AE, puisque la séparation est effectuée sur les distributions *a posteriori* des images spectrales dans le cadre bayésien. De plus, comme on vient de l'expliquer dans la section 5.2, l'AE suppose que la distribution *a posteriori* du mélange admet une décomposition en plusieurs distributions paramétriques et que le choix judicieux des formes paramétriques des distributions influe sur la forme ainsi que la convergence de la fonction de coût d'apprentissage à minimiser de l'algorithme de l'ACI. De ce fait, et afin de rendre l'application de l'ACI plus cohérente avec la nature des modèles probabilistes des images optiques, nous pouvons adopter le modèle gaussien pour les différentes formes paramétriques des distributions qui entrent dans la construction de la fonction de coût de l'ACI.

Considérons, dans ce qui suit, que les informations inconnues (y), sachant le mélange mesuré X, sont les variables inconnues S et θ. Le vecteur θ représente tous les paramètres, y compris la matrice de séparation et le vecteur des sources à estimer (W et S). Soit, le modèle de séparation de sources suivant :

$$X = f(S) + N \tag{10}$$

où N représente le vecteur de bruit supposé gaussien additif, indépendant des sources et ayant une matrice de covariance diagonale. Notons que cette hypothèse sur le bruit est justifiée dans le cas d'images multibandes optiques (Chitroub, 2003). En général, la fonction de la transformation *f* est supposée être la fonction tangente hyperbolique (Lappalainen and Honkela, 2000).

Soit $P(S,\theta|X)$ et $Q(S,\theta|X)$, respectivement, la densité exacte des variables inconnues et son approximation. La mesure de Kullback-Leibler peut être donc reformulée de la manière suivante :

$$C_{S,\theta}(X) = \int_S\int_\theta \frac{Q(S,\theta|X)}{P(S,\theta|X)} dSd\theta \quad (11)$$

La densité de probabilité des variables inconnues S et θ peut être obtenue en utilisant la loi de Bayes :

$$P(S,\theta|X) = \frac{P(X|S,\theta)P(S|\theta)P(\theta)}{P(X)} \quad (12)$$

Le terme $P(X|S,\theta)$ est aussi obtenu à partir de la formule de Bayes. *P(X)* peut être négligé puisqu'il ne dépend d'aucun paramètre. Si l'on considère que S et θ sont indépendants, on a alors : $Q(S,\theta|X) = Q(S|X)Q(\theta|X)$.

Pour les paramètres θ, une densité gaussienne multidimensionnelle ayant une matrice de covariance diagonale est appliquée. Chaque composante du vecteur θ a alors une moyenne $\overline{\theta}_i$ et une variance $\widetilde{\theta}_i$. Ceci implique que l'approximation est un produit de distributions indépendantes $Q(\theta|X) = \prod_i Q_i(\theta|X)$.

Nous considérons que les mêmes hypothèses restent valables pour le vecteur S et, alors, la densité de probabilité $Q(S|X)$ est traitée de la même manière, c'est-à-dire que nous devrons estimer le vecteur moyenne $\overline{S}$ et le vecteur variance $\widetilde{S}$.

Sachant les paramètres estimés, selon les hypothèses supposées, la densité de probabilité exacte $P(S,\theta|X)$ et son approximation $Q(S,\theta|X)$ sont alors des produits de simples distributions gaussiennes. Ainsi, le calcul de la fonction de coût est considérablement simplifié. En propageant les paramètres estimés sur cette fonction de coût, on réitère le processus plusieurs fois jusqu'à l'obtention de la convergence. Ainsi, d'après Lappalainen et Honkela (2000), la mise à jour de la variance et de la moyenne de la composante θ_i est donnée par les itérations suivantes :

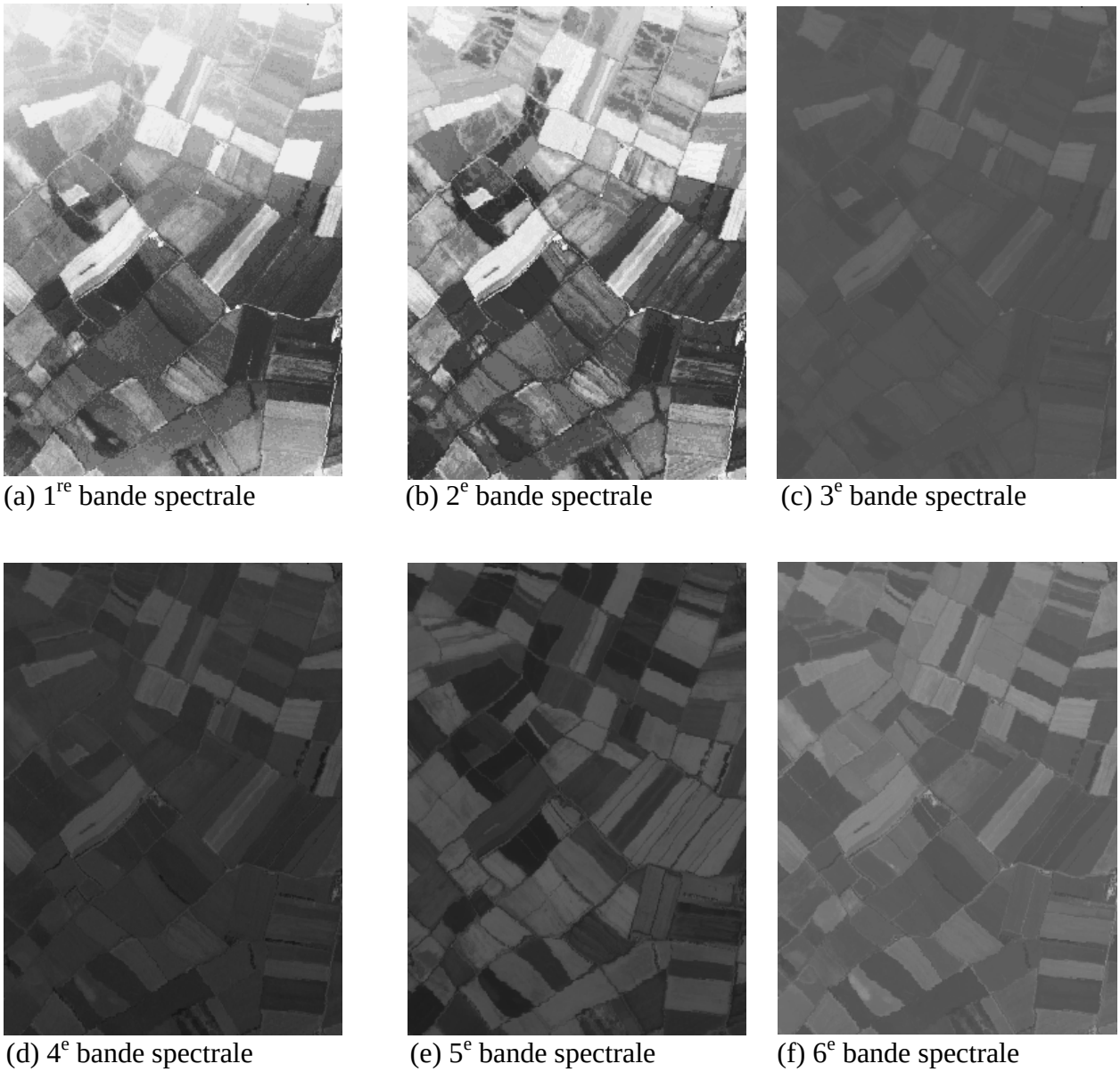
$$\widetilde{\theta}_i = \left(2\frac{\partial C_p}{\partial \widetilde{\theta}_i} \right)^{-1} \quad (13)$$

$$\overline{\theta}_i \leftarrow \overline{\theta}_i - \frac{\partial C_p}{\partial \overline{\theta}_i} \left(\frac{\partial^2 C_p}{\partial \overline{\theta}_i^2} \right)^{-1} \approx \overline{\theta}_i - \frac{\partial C_p}{\partial \overline{\theta}_i} \widetilde{\theta}_i \quad (14)$$

L'algorithme résout les équations (13) et (14) itérativement jusqu'à ce que la convergence soit atteinte. Dans notre application de séparation d'images, une ACP de l'image multibande est effectuée pour trouver les valeurs initiales optimales des $\overline{\theta}_i$. Trouver les valeurs initiales des $\overline{\theta}_i$ par l'ACP est une meilleure solution que d'affecter des valeurs aléatoires aux $\overline{\theta}_i$. En revanche, les valeurs initiales des variances $\widetilde{\theta}_i$ ont des petites valeurs aléatoires.

6. RÉSULTATS EXPÉRIMENTAUX

L'ACI par l'AE pour l'analyse d'image a été testée et évaluée sur une image multibande optique réelle. Une image ATM de six bandes spectrales, de 250 x 350 pixels, représente une région agricole située près de Feltwell, au Royaume-Uni (figure 3). Les champs agricoles sont regroupés en cinq classes correspondant, respectivement, à cinq catégories de cultures : carottes, pommes de terre, sol nu, chaume et betteraves à sucre.



(a) 1re bande spectrale (b) 2e bande spectrale (c) 3e bande spectrale

(d) 4e bande spectrale (e) 5e bande spectrale (f) 6e bande spectrale

Tous droits réservés à Hunting Technical Services Ltd. (Royaume-Uni)

Figure 3 : Image multibande fournie par le système ATM de six bandes spectrales représentant une image de 250 x 350 pixels d'une région agricole située près de Feltwell, au Royaume-Uni. *Multispectral image provided by the six spectral band ATM system representing a scene of 250 x 350 pixels of an agricultural area located near Feltwell in the United Kingdom.*

Afin d'évaluer d'une manière objective et qualitative les performances de l'ACI par l'AE, une étude comparative est effectuée à l'aide de l'algorithme FastICA (Hyvärinen, 1999). Ainsi, les deux algorithmes ont été testés sur la même image multibande. Signalons qu'une version téléchargeable de l'algorithme FastICA se trouve sur le site Internet de l'Université d'Helsinki (www.cs.helsinki.fi/u/ahyvarin/). Cependant, cette version a été modifiée et adaptée au contexte de l'application présentée dans cet article pour réaliser le test effectué sur l'image multibande optique. Des améliorations ont alors été introduites afin d'obtenir un algorithme de séparation des sources optimal, applicable pour notre problème.

Avant de commenter les résultats obtenus à l'aide des deux méthodes d'ACI, il est préférable d'exposer rapidement le principe de base de l'algorithme FastICA et ceci pour mettre en évidence les différences majeures qui existent entre les concepts théoriques des deux méthodes. Ainsi, les arguments avancés concernant les commentaires sur les résultats obtenus seront justifiés.

6.1. Algorithme FastICA

L'algorithme FastICA est un algorithme très performant de maximisation de la fonction de contraste pour les sources non gaussiennes (Hyvärinen, 1999). Il est basé sur le principe de l'algorithme d'apprentissage itératif de type point fixe (*Fixed point algorithm*).

L'algorithme d'apprentissage du réseau cherche à trouver une direction qui est en fait un vecteur w tel que la projection $w^T x$ maximise l'aspect non gaussien. Pour cela, l'algorithme utilise le théorème du point fixe. Si on considère que chaque sortie du réseau est donnée sous forme d'une fonction de type $g_i(w_i^T x)$, tel que g_i est une fonction scalaire non linéaire, la forme simplifiée de l'algorithme pour l'estimation d'une composante indépendante peut être reformulée de la manière suivante :

- initialiser le vecteur w par de faibles valeurs aléatoires ;

- mettre $w^+ = E\left(x.g\left(w^T x\right)\right) - w.E\left(g'\left(w^T x\right)\right)$;

- mettre $w = w^+ / \left\| w^+ \right\|$;

- si la convergence n'est pas encore atteinte, refaire les étapes 2 et 3.

La convergence signifie que les valeurs de w (anciennes et nouvelles) vont dans la même direction, c'est-à-dire que la différence entre ces deux valeurs est au-dessous du critère de convergence entre deux itérations.

Pour estimer plusieurs composantes, on a besoin d'utiliser l'algorithme précédent avec plusieurs neurones ayant les vecteurs synaptiques $w_1, w_2, ...w_n$. Pour que les vecteurs ne convergent pas vers un même maximum, nous devons décorréler les projections $w_1^T x, w_2^T x, ..., w_n^T x$ après chaque itération. Il existe plusieurs méthodes de décorrélation. La plus populaire est basée sur la méthode de Gram-Shmidt de décomposition de matrices. Avec cette méthode, après l'estimation de p composantes indépendantes $\left(w_1, w_2, ...w_p\right)$, on utilise l'algorithme du point fixe pour estimer le vecteur w_{p+1} et, après chaque itération, on soustrait du vecteur w_{p+1} les projections $w_{p+1}^T w_j w_j$, $\left(j = 1, ...p\right)$, puis on normalise de nouveau le vecteur w_{p+1} :

- mettre $w_{p+1} = w_{p+1} - \sum_{j=1}^{p} w_{p+1}^T w_j w_j$;

- mettre $w_{p+1} = w_{p+1} / \sqrt{w_{p+1}^T w_{p+1}}$.

La performance de l'algorithme FastICA dépend du choix de la fonction g. Les composantes indépendantes peuvent être extraites une à une. Ainsi, on obtient un gain considérable en performance si l'on a seulement besoin de quelques composantes indépendantes à extraire.

6.2. Commentaires

Tout d'abord, nous constatons que la méthode de la FastICA est plus rapide par rapport à la méthode de l'ACI par l'AE. Sur le tableau 1, on récapitule les performances en termes de temps de calcul des deux algorithmes (sur un ordinateur doté d'un processeur Pentium 4E, 2.8Ghz). Notons que les seuils de convergence ont été fixés à 10^{-3} pour la FastICA et à $5{,}10^{-4}$ pour l'ACI par l'AE. Ces seuils ont été choisis de façon empirique. Cette différence dans la convergence des deux algorithmes peut être justifiée par le fait que, dans l'algorithme de l'ACI par l'AE et pour chaque itération, le nombre de paramètres qui doivent être estimés est plus élevé que le nombre de paramètres nécessaires pour la mise à jour de l'algorithme FastICA. Par contre l'ACI par l'AE est plus performante que la méthode FastICA en termes des images composantes indépendantes obtenues.

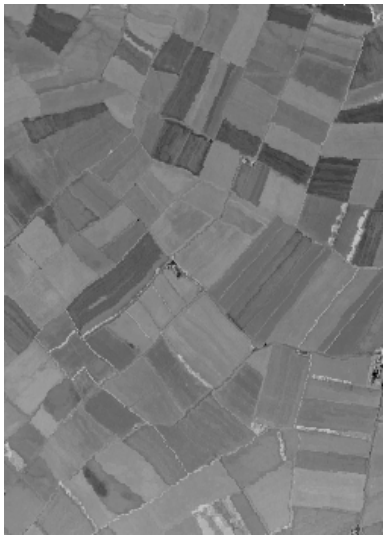
Tableau 1 : Les performances en termes de temps de calcul des algorithmes ACI par AE et FastICA sur un ordinateur doté d'un processeur Pentium 4E, 2,8Ghz. *Performances in terms of computing time of the ACI by ensemble learning and FastICA algorithms respectively (on a machine equipped with a Pentium processor 4E, 2.8 GHz).*

Algorithme	Temps de calcul en secondes	Nombre d'itérations
FastICA	61	6 062
ACI par AE	3 661	354

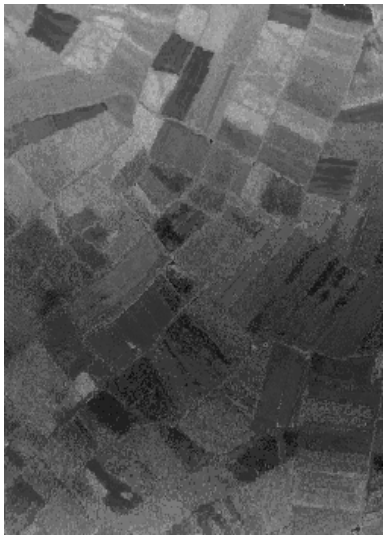
Sur les figures 4 et 5, on montre les résultats obtenus à l'aide de l'ACI par l'AE et de la FastICA. En effet, l'algorithme FastICA est essentiellement basé sur une application neuronale des statistiques d'ordre supérieur (kurtose d'ordre trois et quatre) des données d'entrée au moyen de la fonction de contraste des sources non gaussiennes. Or, les images optiques de télédétection sont parfaitement modélisées par le modèle gaussien mono et multidimensionnel (Chitroub, 2003). Le modèle gaussien est caractérisé par les statistiques d'ordre un et deux, c'est-à-dire la moyenne et la matrice de covariance. Les statistiques d'ordre supérieur à deux d'un modèle gaussien sont toutes nulles (Hyvärinen *et al*., 2001). Ainsi, l'incohérence entre les hypothèses de départ de la FastICA et la nature du modèle statistique de l'image multibande est une cause de la dégradation des résultats obtenus. En revanche, pour l'ACI par l'AE, l'approche bayésienne sur laquelle la méthode est basée a allégé la contrainte de modélisation statistique des données d'entrée. Les modèles gaussiens ont été considérés pour les paramètres des distributions *a posteriori* constituant le mélange d'entrée (dans notre cas, il s'agit des distributions des images spectrales d'entrée) ainsi que pour les sources à séparer. De ce fait, l'ACI par l'AE est devenue plus cohérente avec la nature probabiliste de l'image multibande d'entrée. Ceci est justifié par les images composantes indépendantes obtenues qui sont plus claires et plus contrastées que celles obtenues à partir de la FastICA. Cependant, nous constatons que l'ordre dans lequel les images composantes indépendantes ont été produites pour les deux algorithmes est aléatoire et non justifié. Par exemple, la meilleure composante obtenue avec ACI par l'AE n'était pas la première mais plutôt la sixième. Ceci est dû au fait que l'initialisation des matrices de projection des deux algorithmes est aléatoire. Nous pouvons, cependant, améliorer les résultats de l'ACI par l'AE en introduisant le concept d'extraction orientée des composantes indépendantes par un choix étudié de la matrice de projection initiale.



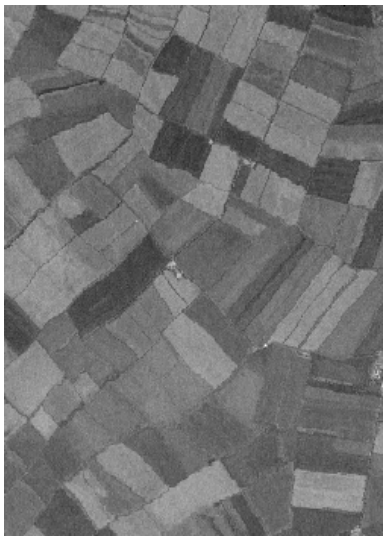
(a) 1re image composante



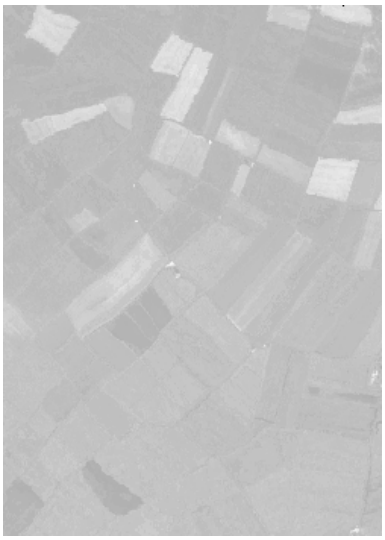
(b) 2e image composante



(c) 3e image composante



(d) 4e image composante

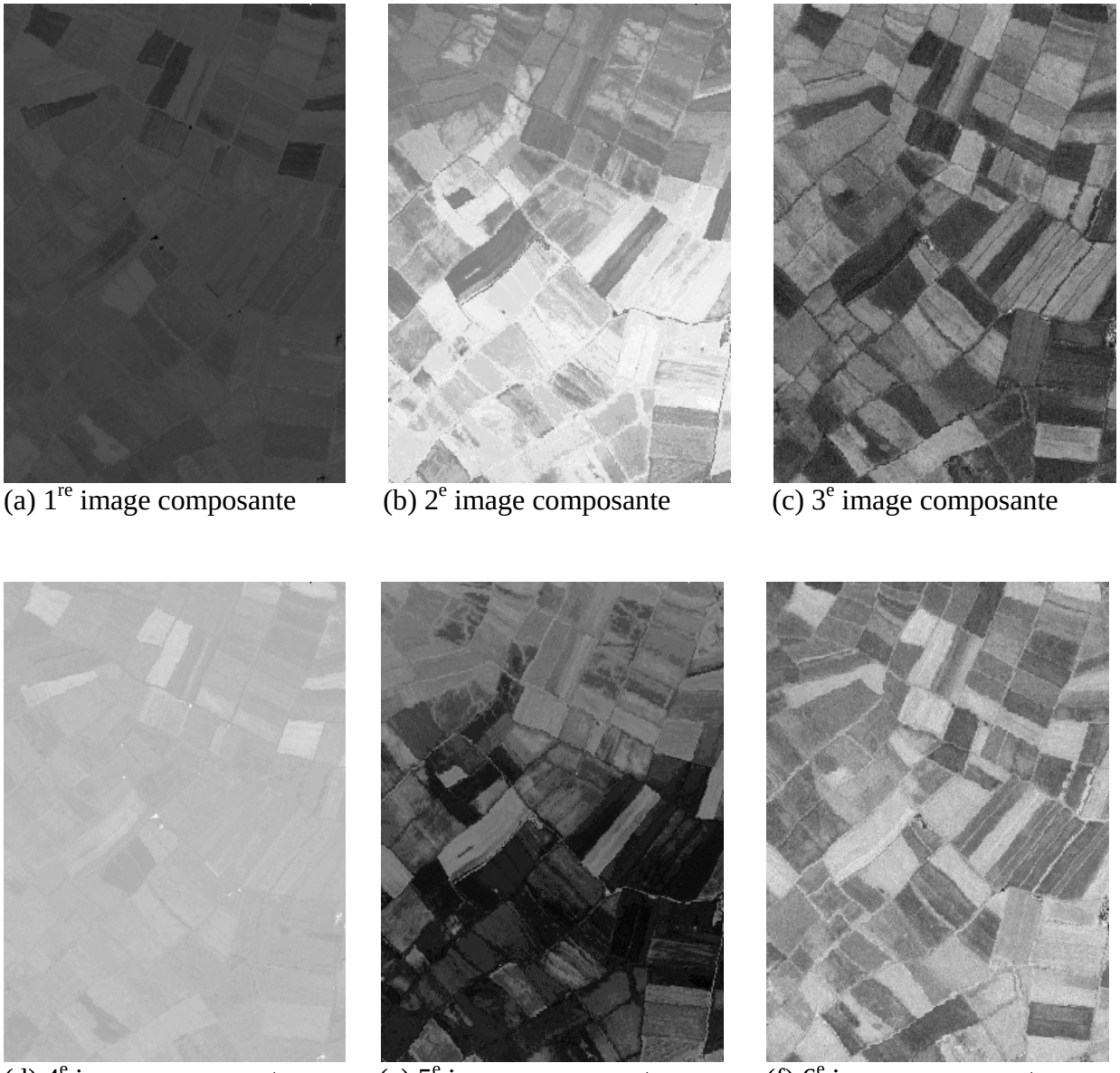


(e) 5e image composante



(f) 6e image composante

Figure 4 : Les six images de composantes indépendantes résultant de l'application de l'ACI par AE sur l'image multibande.*The six independent component images produced by applying the ACI by ensemble learning on the multispectral image.*



(a) 1re image composante (b) 2e image composante (c) 3e image composante

(d) 4e image composante (e) 5e image composante (f) 6e image composante

Figure 5 : Les six images de composantes indépendantes résultant de l'application de l'algorithme FastICA sur l'image multibande. *The six independent component images produced by applying the FastICA algorithm on the multispectral image.*

7. CONCLUSION

Dans cet article, l'ACI en télédétection, sa faisabilité et ses perspectives ont été abordées. Ainsi, les problèmes de cohérence entre les fondements théoriques de l'ACI et son application pour l'analyse d'images multibandes optiques ont été mis en évidence. Les perspectives concernant son application pour la réduction de la dimension et l'extraction des membres extrêmes ont été abordées. Dans ce contexte, l'ACI par l'AE pour l'analyse d'images multibandes optiques est proposée. Sa mise en œuvre est réalisée. Bien que l'application de son algorithme soit fastidieuse (plusieurs hypothèses sont à vérifier et plusieurs paramètres sont à estimer pour chaque itération), la force de l'ACI par l'AE réside dans le fait que la contrainte de modélisation probabiliste des sources est allégée puisque le principe bayésien constitue la base de l'AE. Ainsi, l'application de l'ACI sur de l'imagerie multibande est devenue plus cohérente. Les résultats obtenus sont en général acceptables. L'étude comparative entre l'ACI par l'AE et la FastICA a montré que l'ACI par l'AE présente une certaine supériorité en termes de qualité des images composantes indépendantes extraites. Cependant, l'inconvénient majeur de l'ACI par l'AE est le temps de calcul qui est prohibitif. Ceci est dû principalement au nombre de paramètres à calculer durant l'apprentissage, mais les résultats obtenus justifient sa lourdeur puisqu'en fait l'ACI recherche un minimum global et non un minimum local. Un autre inconvénient de l'ACI par l'AE est l'ordre aléatoire dans lequel les composantes indépendantes sont extraites. Cet inconvénient pose le problème de l'utilisation des composantes indépendantes extraites pour l'amélioration de la classification de la couverture terrestre. Néanmoins, cet inconvénient pourrait être pallié en utilisant l'extraction orientée des composantes. Dans le futur, nous essaierons

d'améliorer l'ACI par l'AE afin de rendre la mise en œuvre de son algorithme plus simple et, ainsi, réduire le temps de calcul tout en préservant sa précision dans le processus d'extraction des images composantes indépendantes.

REMERCIEMENTS

Cet article est dédié à la mémoire de Ferdinand Bonn. Quand j'ai commencé à préparer ma thèse de magister, il y a dix ans, dans un domaine tout a fait nouveau pour moi, soit le traitement d'images de télédétection, les choses n'étaient pas claires et j'avais besoin d'un repère pour me positionner et démarrer ma recherche. Heureusement, j'avais trouvé le manuel « Précis de télédétection : principes et méthodes » qui m'avait guidé à tracer mon chemin en toute simplicité et sans ambiguïtés. À cet effet, je ne saurais terminer sans mentionner tout ce que je dois à Ferdinand Bonn. Enfin, je remercie les professeurs Sebastiano B. Serpico et Farid Melgani, de l'Université de Gênes, en Italie, de nous avoir fourni l'image multibande optique acquise par le système d'acquisition d'images optiques ATM à six bandes spectrales.

RÉFÉRENCES

Achard, S. (2003) Mesures de dépendance pour la séparation aveugle de sources. Application aux mélanges post non linéaires. Thèse de doctorat, Université de Grenoble, Grenoble, 191 p.

Bayliss, J., Gualtieri, A. and Cromp, R. (1998) Analysing hyperspectral data with independent component analysis. P. 133-143, *in* J.M. Selander (réd.) Proceedings of the 26th Artificial Intelligence and Pattern Recognition (AIPR) Workshop, Washington (D.C.), 15-17 octobre 1997, SPIE, Bellingham (WA), vol. 3240.

Bonn, F. et Rochon, G. (1992) Précis de télédétection - volume 1 : Principes et méthodes. AUF, Presses de l'Université du Québec et AUPELF, Sainte-Foy et Montréal, 485 p.

Botchko, V., Berina, E., Korotkaya, Z., Parkkinen, J. and Jaaskelainen, T. (2003) Independent component analysis in spectral images. P. 203-207, *in* S. Amari, A. Cichocki, S. Makino and N. Murata (réd.) Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003), Nara (Japon), 1-4 avril 2003, NTT Communication Science Laboratories, Kyoto, 1120 p.

Campbell, J.B. (1996) Introduction to Remote Sensing. 2e édition, The Guilford Press (Taylor and Francis), London, 622 p.

Cardoso, J.F. (1999) High-order contrasts for independent component analysis. Neural Computation, vol. 11, n° 1, p. 157-192.

Chang, C.-I. and Du, Q. (2004) Estimation of number of spectrally distinct signal sources in hyperspectral imagery. IEEE Transactions on Geosciences and Remote Sensing, vol. 42, no 3, p. 608-619.

Chang, C.-I. and Plaza, A. (2006) Fast iterative algorithm for implementation of pixel purity index. Geoscience and Remote Sensing Letters, vol. 3, n° 1, p. 63-67.

Chang, C.-I., Chiang, S.S., Smith, J.A. and Ginsberg, I.W. (2002) Linear spectral random mixture analysis for hyperspectral imagery. IEEE Transactions on Geosciences and Remote Sensing, vol. 40, n° 2, p. 375-392.

Chitroub, S. (2003) Modélisation, analyse, classification et fusion d'images de télédétection multisources optiques et radars. Thèse de doctorat, Faculté d'électronique et d'informatique, Université des sciences et de la technologie Houari Boumediene, Alger, n° 01/2003-E/EL, 245 p.

Chitroub, S. (2004) PCA-ICA neural network model for POLSAR images analysis. P. 757-760, *in* Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing (ICASSP'04), Montréal, 17-21 mai 2004, IEEE, Piscataway (N.J.).

Chitroub, S. and Hachemi R. (2006) Independent component analysis by ensemble learning applied to polarimetric synthetic radar image analysis. P. 76-80, *in* N. Kerle and A. Skidmore (réd.) Proceedings of ISPRS Commission VII Mid-term Symposium "Remote Sensing : From Pixels to Processes", Enschede (Pays-Bas), 8-11 mai 2006, ISPRS, 800 p.

Chitroub, S., Houacine, A. and Sansal, B. (2000) Compound PCA-ICA neural network model for enhancemat and feature extraction of multi-frequency polarimetric SAR imagery. P. 328-331, *in* Proceedings of IEEE International Conference on Image Processing (ICIP'00), Vancouver, 10-13 septembre 2000, IEEE, Piscataway (N.J.), vol. 3, 1000 p.

Chitroub, S., Houacine, A. and Sansal, B. (2002a) Statistical characterisation and modelling of SAR images. Signal Processing, vol. 82, no 1, p. 69-92.

Chitroub, S., Houacine, A. and Sansal, B. (2002b) Unsupervised learning rules for POLSAR images analysis. P. 567-576, *in* Proceedings of 12th IEEE Workshop on Neural Network for Signal Processing, Martigny (Suisse), 3-6 septembre 2002, IEEE, Piscataway (N.J.), 790 p.

Cikochi, A. and Amari, S. (2002) Adaptive Blind Signal and Image Processing. John Wiley & Sons, New York, 555 p.

Comon, P. (1994) Independent component analysis, a new concept ? Signal Processing, vol. 63, n° 3, p. 287-314.

Cover, T. and Thomas, J. (1991) Elements of Information Theory. John Wiley & Sons, New York, vol. 1, 527 p.

Darmois, G. (1953) Analyse générale des liaisons stochastiques. Revue de l'Institut international de statistique, vol. 21, p. 2-8.

Green, A.A., Berman, M., Switzer, P. and Craig, M.D. (1988) A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geosciences Remote and Sensing, vol. 26, n° 1, p. 65-74.

Hughes, G. F. (1968) On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory, vol. 14, n° 1, p. 55-63.

Hyvärinen, A. (1999) Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks, vol. 10, no 3, p. 626-634.

Hyvärinen, A., Karhunen, J. and Oja, E. (2001) Independent Component Analysis. John Wiley & Sons, New York, 481 p.

Jutten, C. (1987) Calcul neuromimétique et traitement du signal, analyse en composantes indépendantes. Thèse de doctorat, Université de Grenoble, Grenoble, 200 p.

Kullback S. and Leibler, R. A. (1951) On information and sufficiency. The Annals of Mathematical Statistics, vol. 22, p. 79-86.

Landgrebe, D. (1999) P. 3-33 *in* C. H. Chen (réd.) Information Processing for Remote Sensing. Chapter 1 : Information Extraction Principles and Methods for Multispectral and Hyperspectral Image Data. World Scientific Publishing, River Edge (N.J.).

Lappalainen, H. and Honkela, A. (2000) Bayesian non-linear independent component analysis by multilayer perceptrons. P. 93-121, *in* M. Girolami (réd.) Advances in Independent Component Analysis. Springer-Verlag, Berlin, 93 p.

Lappalainen, H. and Miskin, J. (2000) Ensemble learning. P. 75-92, *in* M. Girolami (réd.) Advances in Independent Component Analysis. Springer-Verlag, Berlin, 279 p.

Lee, T.W., Girolami, M. and Sejnowski, T.J. (1999) Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Computation, vol. 11, p. 417-441.

Lee, T.W., Girolami, M., Bell, A.J. and Sejnowski, T.J. (2000) A unifying information-theoretic framework for independent component analysis. Neural Networks, vol. 39, p. 1-21.

Lennon, M. (2002) Méthodes d'analyse d'images hyperspectrales. Exploitation du capteur aéroporté CASI pour des applications de cartographie agro-environnementale en Bretagne. Thèse de doctorat, Université de Rennes I, Rennes, n° 2780, 345 p.

Loghmari, M.A., Naceur M.S. and Boussema, M.R. (2006) A spectral and spatial source separation of multispectral images. IEEE Transactions on Geosciences and Remote Sensing, vol. 44, n° 12, p. 3659-3673.

Mackay, D.J.C. (1995) Developments in probabilistic modeling with neural networks – ensemble learning. P. 191-198, *in* B. Kappen and S. Gielen (réd.) Neural Networks: Artificial Intelligence and Industrial Applications. Proceedings of the 3rd Annual Symposium on Neural Networks, Nijmegen (Pays-Bas), 14-15 septembre 1995, Springer, Berlin, 191 p.

Nascimento, J.M.P. and Bioucas Dias, J.M. (2005) Does independent component analysis play a role in unmixing hyperspectral data?. IEEE Transactions on Geosciences and Remote Sensing, vol. 43, n° 1, p. 175-187.

Neal, R.M. (1996) Bayesian Learning for Neural Networks. Springer-Verlag, Berlin, Lecture Notes in Statistics no 118, 204 p.

Parra, L., Mueller, K.R., Spence, C., Ziehe, A. and Sajda, P. (2000) Unmixing hyperspectral data. Advances in Neural Information Processing Systems, vol. 12, p. 942–948.

Ramakishna, B., Plaza, A., Chang, C.-I., Ren, H., Du, Q. and Chang, C.-C. (2005) Spectral/spatial hyperspectral image compression. P. 309-346, chap. 11, *in* G. Motta, F. Rizzo and J. Storer (réd.) Hyperspectral data compression, Springer-Verlag, New York, 426 p.

Ristaniemi, T. and Joutsensalo, J. (1999) On the performance of blind source separation in CDMA downlink. P. 437-441, *in* Proceedings of the International Workshop on Independent Component Analysis and Signal Separation (ICA'99), Aussois (France), 11-15 janvier 1999, IEEE, Piscataway (N.J.), 437 p.

Settle, J.J. and Drake, N.A. (1993) Linear mixing and the estimation of ground cover proportion. International Journal of Remote Sensing, vol. 14, n° 6, p. 1159-1177.

Shannon, C.E. (1948) A mathematical theory of communication. Bell System Technical Journal, vol. 27, p. 379-423 et 623-656.

Wang, J. and Chang, C.-I. (2006a) Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. IEEE Transactions on Geosciences and Remote Sensing, vol. 44, n° 6, p. 1586-1600.

Wang, J. and Chang, C.-I. (2006b) Applications of independent component analysis in end-member extraction and abundance quantification for hyperspectral imagery. IEEE Transactions on Geosciences and Remote Sensing, vol. 44, n° 9, p. 2601-2616.

Winter, M.E. (1999) N-FINDER : an algorithm for fast autonomous spectral end-member determination in hyperspectral data. P. 266-275, *in* R.M. Descour and S.S. Shen (réd.) Proceedings of SPIE Conference on Imaging Spectrometry V, Denver, 27 octobre 1999, SPIE publication no 3753, Bellingham (W.A.), 266 p.

Zhang, X. and Chen, C.H. (2002) New independent component analysis method using higher order statistics with applications to remote sensing images. Optical Engineering, vol. 41, p. 1717–1728.