

Approches neuronales pour l'extraction des composantes principales d'images multispectrales de télédétection

Salim Chitroub, Amrane Houacine, Abdel-Azziz Khadraoui et Boualem Sansal
Laboratoire de Traitement du Signal, Institut d'Electronique

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

B. P. 32, El-Alia, Bab-Ezzouar, Alger, ALGERIE

Fax : (213) (2) 24-71-87, E-mail : signal@ist.cerist.dz

Résumé - Le problème traité dans le présent article consiste en l'extraction des composantes principales les plus significatives d'images multispectrales de télédétection sans avoir à calculer la matrice de covariance des images spectrales. L'originalité du travail réside dans l'élaboration des algorithmes d'apprentissage spécifiques pour deux approches neuronales d'Analyse en Composantes Principales (ACP). Les deux approches possèdent des convergences rapides. L'application sur une image multispectrale réelle a montré leur efficacité dans l'extraction des composantes principales les plus significatives.

Abstract - The problem addressed in the present paper is the most significant principal components extraction of remotely sensed multispectral images without having to calculate the covariance matrix of spectral images. The originality of the work resides in the elaboration of specific training algorithms for two neural network-based approaches of Principal Component Analysis (PCA). The convergence of a proposed approaches are rapid. The application on a real multispectral image has shown their efficiency in the extraction of the most significant principal components.

1. Introduction

Le traitement d'images de télédétection constitue aujourd'hui un moyen privilégié de prévision, de gestion et de décision pour tout ce qui concerne les ressources terrestres. En effet, une image multispectrale fournit la solution économique pour disposer d'informations récentes sur toute zone d'intérêt. Cependant, celle-ci n'est pas exploitable directement. La corrélation entre les images spectrales, due à la proximité spectrale des capteurs, complique les traitements à effectuer. Le processus de formation de ces images est également soumis à des perturbations diverses considérées comme un bruit additif [1]. Une analyse descriptive de l'image multispectrale est alors nécessaire. Le but est d'obtenir une représentation optimale de l'image multispectrale dans laquelle l'information portée est concentrée sur un nombre réduit de nouvelles images décrivant la scène observée. L'Analyse en Composantes Principales (ACP) a été utilisée afin d'obtenir un nombre restreint d'images composantes principales (CP) décorréliées dont la variance est maximale et le rapport signal sur bruit (SNR) est amélioré. L'application de l'ACP consiste à calculer la matrice de covariance des images spectrales, Σ_X , puis la diagonaliser, en calculant les valeurs propres et les vecteurs propres associés. Les valeurs propres présentent les variances des CP. Les vecteurs propres associés sont les coefficients des combinaisons linéaires des CP [1][2]. Si la taille et le nombre d'images spectrales augmentent, la dimension de la matrice augmente aussi. Ainsi, l'accumulation de l'erreur dans le calcul de Σ_X devient importante et l'évaluation précise de ses valeurs et vecteurs propres n'est plus possible. De plus, l'ACP ne peut être appliquée si on ne fixe pas a priori le nombre d'images spectrales d'entrée.

Nous proposons dans le présent article deux nouvelles approches de l'ACP en utilisant les réseaux de neurones. Les deux approches, basées sur le calcul neuronal adaptatif et récursif, permettent l'extraction des CP les plus significatives sans avoir besoin de calculer la matrice Σ_X . Les analyses théoriques des deux approches sont fondées sur les règles d'apprentissage d'Hebbian normalisées [3][4][5] et sur la théorie de la transformation de déflation de matrices [6]. Les

performances des approches sont optimisées en spécifiant les algorithmes d'apprentissage qui prendront en considération de façon implicite les critères d'optimisation des réseaux.

Le reste de cet article est organisé comme suit. Dans le paragraphe 2, la transformation de déflation de matrices est donnée. On décrit dans le paragraphe 3 les règles d'apprentissage d'Hebbian normalisées. Le paragraphe suivant est consacré à la description des deux approches neuronales de l'ACP proposées. Les résultats obtenus sur une image multispectrale réelle fournie par le satellite Landsat-TM sont exposés et commentés dans le paragraphe 5. Enfin, la conclusion est donnée au dernier paragraphe.

2. Transformation de déflation de matrices

La recherche de valeurs propres et vecteurs propres de matrices est un problème assez difficile dans le cas général. Cependant, en ACP on a affaire à une matrice de covariance qui est une matrice carrée, réelle, symétrique et définie positive. Pour de telles matrices le problème est considérablement simplifié du fait que les vecteurs propres correspondants à des valeurs propres distinctes sont orthogonaux. Plusieurs méthodes de recherche de valeurs propres et vecteurs propres ont été développées dans la littérature [6]. On s'intéresse ici à la méthode de calcul de valeurs propres et vecteurs propres par la transformation de déflation de matrices. Le principe de la méthode est le suivant : si A est une matrice carrée de dimension n avec comme valeurs propres $\lambda_1 > \lambda_2 > \dots > \lambda_n > 0$ et vecteurs propres d_1, d_2, \dots, d_n , alors la matrice A_1 , telle que :

$$A_1 = (I - d_1 d_1^t) A \quad (1)$$

admet comme valeurs propres $0, \lambda_2, \dots, \lambda_n$ et vecteurs propres d_1, d_2, \dots, d_n . L'expression (1) est appelée la transformation de déflation de la matrice A . Il suffit donc d'appliquer la méthode de la puissance itérée [6] à A_1 pour obtenir la valeur propre de A de plus grand module après λ_1 , puis d'opérer une déflation de A_1 , et ainsi de suite. Cependant, la méthode de la puissance itérée sur A ne donne pas λ_1 et d_1 , mais des valeurs approchées. Ainsi, par déflation, on n'obtient pas la matrice A_1

mais une matrice approchée. Il y a dégradation de la matrice de départ, qui s'ajoute à l'approximation de la méthode proprement dite. La précision sur les valeurs propres et les vecteurs propres successifs diminue à partir d'un certain rang de la matrice.

3. Règles d'Hebbian normalisées

Au cours de ces dernières années, des développements importants dans la théorie des réseaux de neurones et de calcul neuronal ont été réalisés. Les règles d'apprentissage d'Hebbian [3] consiste en la mise à jour, proportionnellement au produit entrées-sortie correspondant, de chaque coefficient synaptique connectant les entrées au neurone de sortie. Le neurone s'organise dans le sens de reconnaître les caractéristiques spécifiques des entrées. Ce comportement a un lien avec le critère statistique de l'ACP lorsque le modèle de neurone est linéaire. En effet, en 1982 Oja [3] a montré que la version normalisée des règles d'Hebbian appliquées sur un simple neurone linéaire converge vers le vecteur propre de la matrice de corrélation des variables d'entrées. La règle d'Hebbian est numériquement instable [3][4]. L'algorithme de normalisation proposé par Oja dans [4] pour stabiliser les règles d'Hebbian peut être formulé par l'équation itérative suivante :

$$M(k) - M(k-1) = \beta(x(k) - M(k-1)y(k)), y^t(k) \quad (2)$$

où $x(k)$ est le processus stochastique d'entrée, $y(k) = M^t(k).x(k)$ est la sortie du neurone et M le vecteur des coefficients synaptiques. β est un paramètre positif compris entre 0 et 1. Le terme $(x(k).y^t(k))$ correspond à la règle d'Hebbian standard et $(M(k-1).y(k).y(k)^t)$ le terme utilisé pour stabiliser le réseau. L'équation différentielle ordinaire correspondant à l'expression (2) est :

$$dM(k)/dk = C.M(k) - M(k).M^t(k).C.M(k) \quad (3)$$

où C est la matrice de covariance du processus d'entrée admettant la décomposition spectrale $C = D.A.D^t$. Soit le vecteur $w(k) = D^t.M(k)$, l'expression (3) devient alors :

$$dw(k)/dk = (I - w(k).w^t(k)).A.w(k) \quad (4)$$

L'équation (4) est similaire à l'équation (1). Ceci montre qu'un réseau de neurones linéaire converge vers le vecteur propre de la matrice de covariance des variables d'entrées. Depuis l'article d'Oja (1982), plusieurs approches basées sur les réseaux de neurones ont été proposées pour l'extraction d'une ou plusieurs CP d'un processus stochastique stationnaire directement à partir des variables d'entrée [7][8][9][10].

4. Approches neuronales proposées

Les deux approches neuronales de l'ACP que nous présentons ici sont basées sur les règles d'Hebbian normalisées et le principe de calcul de valeurs propres et vecteurs propres par la méthode de la transformation de déflation de matrices. L'optimisation des deux approches est implicite dans leurs algorithmes d'apprentissage [11].

4. 1. Première approche neuronale de l'ACP

Le modèle du réseau est présenté en (Fig.1). L'entrée est l'image multispectrale, X , et les sorties sont les images CP, soient Y_m . L'extraction des CP se fait de façon séquentielle composante après composante. W_i est le vecteur des coefficients synaptiques connectant les entrées à chaque neurone de sortie. C_i est le vecteur des coefficients latéraux qui permet d'introduire les $(m-1)$ CP déjà extraites dans la construction de la $m^{ème}$ CP. L'apprentissage se fait selon les deux règles suivantes :

$$W_m^{(k+1)} = W_m^{(k)} + \beta_k.(Y_m.X - Y_m^2.W_m^{(k)}) \quad (5)$$

$$C_m^{(k+1)} = C_m^{(k)} - \beta_k.(Y_m.X + Y_m^2.C_m^{(k)}) \quad (6)$$

La première règle, dite adaptative, est la règle d'apprentissage d'Hebbian normalisée. Elle sert à l'extraction de toutes les CP. La deuxième règle, dite d'orthogonalité et inspirée du principe de déflation de matrices, est la règle d'apprentissage anti-Hebbian. Elle n'est pas utilisée dans la construction de la première CP. Elle permet de rendre la $m^{ème}$ CP orthogonale aux $(m-1)$ CP déjà extraites. Le paramètre β_k joue un rôle capital dans la détermination des performances du réseau. En effet, afin de concevoir un algorithme d'apprentissage récursif optimal, l'évolution de β_k doit être cohérente avec la nature des données et l'application envisagée et en même temps ne doit pas conduire à une convergence lente [8][9]. L'idée est d'intégrer le critère d'optimisation du réseau dans les règles (1) et (2) via un choix optimal de la valeur de β_k . Pour cela, nous avons exprimé β_k en fonction de la taille de l'image, N , et de la variance de la CP en cours de construction, $E(y_m^2)$, telle que :

$$\beta_k = 1/(N.E(y_m^2)) \quad (7)$$

La mise en œuvre de cette idée a nécessité quelques considérations pratiques afin de rendre le déroulement de l'algorithme d'apprentissage simple et plus rapide. Le réseau se stabilise lorsque les vecteurs W_i convergent vers les vecteurs propres de la matrice Σ_x , et les vecteurs C_i vers zéro.

4. 2. Deuxième approche neuronale de l'ACP

Le modèle du réseau est présenté en (Fig. 2). Les entrées sont les images spectrales et les sorties sont les images CP. L'extraction des CP se fait d'une manière séquentielle. La première CP est extraite en minimisant l'erreur des moindres carrés suivante :

$$J_I(n) = \sum_{k=1}^n (X - Y_I^{(k)}.W_I^{(k-1)})^t.(X - Y_I^{(k)}.W_I^{(k-1)}) \quad (8)$$

$$\text{avec : } Y_I^{(k)} = (W_I^{(k-1)})^t.X \quad (9)$$

Le vecteur W_I est le vecteur des coefficients synaptiques connectant les neurones d'entrée au premier neurone de sortie et n est le nombre d'itérations. La mise à jour du vecteur W_I est effectuée selon la règle d'apprentissage d'Hebbian normalisée telle que :

$$W_I^{(k)} = W_I^{(k-1)} + \beta_I^{(k)} \cdot (X - Y_I^{(k)} \cdot W_I^{(k-1)}) Y_I^{(k)} \quad (10)$$

$$\text{avec: } \beta_I^{(k)} = \left(D_I^{(k-1)} \cdot Y_I^{(k)} \right) / \left(1 + \left(Y_I^{(k)} \right)^2 \cdot D_I^{(k-1)} \right) \quad (11)$$

$$\text{et: } D_I^{(k)} = \left(1 - \beta_I^{(k)} \cdot Y_I^{(k)} \right) \cdot D_I^{(k-1)} \quad (12)$$

L'algorithme d'apprentissage pour l'extraction des autres CP est basé sur le principe de calcul des valeurs propres et des vecteurs propres en utilisant la transformation de déflation de matrices et en tenant compte du critère d'optimisation des moindres carrés [10]. Les entrées du réseau ne sont plus les images spectrales mais des entrées calculées par déflation. En effet, le critère d'optimisation (3) est une fonction convexe par rapport au vecteur W_I , elle possède donc un minimum global unique. Ce minimum est atteint lorsque le vecteur W_I calculé correspond au vecteur propre de la vraie matrice de covariance Σ_X [9]. Par conséquent, en utilisant de façon récursive les entrées calculées par déflation comme entrée du réseau, les équations (3)-(7) peuvent être généralisées pour l'extraction du reste des CP. La stabilité du réseau est alors caractérisée par la convergence des vecteurs W_i vers les vecteurs propres de la vraie matrice Σ_X . Les variances des CP sont déterminées en divisant la valeur finale de D_i par le nombre total d'itérations.

5. Résultats expérimentaux

L'image multispectrale (6 images spectrales) utilisée pour tester et évaluer les approches neuronales proposées est une image de taille (200x200pixels) provenant du satellite Landsat-TM et couvrant la région Laghouat City dans le Sud de l'Algérie. Selon le tableau (Tab. 1), la deuxième image spectrale possède le SNR le plus faible (image.1) et la cinquième image spectrale le SNR le plus fort (image.2). Les premières CP obtenues par les approches neuronales sont très nettes et similaires (images 3 et 4). L'amélioration du SNR dans ces CP par rapport aux images originales est importante (Tab. 2). L'information portée par ces CP caractérise d'une manière plus précise l'état de surface de la scène observée. Les deuxièmes CP (images 5 et 6) sont moins contrastées et le SNR est faible (Tab. 2). Le reste des CP est dominé par le bruit et l'information portée est négligeable. Les figures 3 et 4 nous donnent l'évaluation des valeurs propres obtenues par les deux approches neuronales et par la méthode de déflation de matrices [6]. Nous remarquons que les valeurs propres les plus significatives fournis par les trois algorithmes sont très proches. La même remarque peut être formulée pour les vecteurs propres associés (Fig.5 et Fig.6). Les convergences des deux approches sont données par les figures (Fig.7) et (Fig.8), respectivement. Les convergences pour l'extraction des premières CP sont moins rapides par rapport à celles relatives à l'extraction du reste des CP. Ceci est justifié en se référant aux analyses théoriques des deux approches.

6. Conclusions

Le but essentiel du travail présenté ici était l'extraction des CP de l'image multispectrale de télédétection sans avoir à calculer la matrice de covariance des images spectrales originales. Les réseaux de neurones permettent de réaliser cette idée, mais l'inconvénient majeur reste le temps d'apprentissage qui est généralement important. Cependant, les deux approches neuronales de l'ACP proposées ici ne présentent pas cet inconvénient. Ceci est dû aux algorithmes d'apprentissage spécifiques élaborés qui tiennent compte d'une manière implicite des critères d'optimisation des réseaux. Ces critères sont établis à l'analyse théorique et en tenant compte de la nature des données, de l'application envisagée et du critère d'arrêt des itérations. Les résultats obtenus montrent l'efficacité des deux approches dans l'extraction des CP les plus significatives de l'image multispectrale de télédétection.

Références

- [1] J. A. Richards, *Remote Sensing and Image Analysis*, Springer-Verlag, Berlin, 1986, Chap. 1, PP. 5-25, Chap. 10, PP. 206-224.
- [2] W. P. Laughlin, "Principal component analysis for alteration mapping", *Photogrammetric Engineering & Remote Sensing*, Vol. 57, N°9, Sept. 1991, PP. 1165-1169.
- [3] E. Oja, "A simplified neuron model as a principal component analyser", *J. Math. Biology*, Vol. 15, 1982, PP. 267-273.
- [4] E. Oja, "Neural networks, principal components, and subspaces", *Int. J. of Neural Systems*, Vol. 1, N°1, 1989, PP. 61-68.
- [5] T. Chen, Y. Hua and W. Y. Yan, "Global convergence of Oja's subspace algorithm for principal component extraction", *IEEE Trans. on Neural Networks*, Vol. 9, N°1, Jan. 1998, PP. 58-66.
- [6] J. P. Nougier, *Méthodes de calcul numériques*, 3^{ème} édition révisée, Edition Masson, 1987, Chap. 7, PP. 148-165.
- [7] J. Mao and A. K. Jain, "Artificial neural networks for feature extraction and multivariate data projection", *IEEE Trans. on Neural Networks*, Vol. 6, N°2, Mar. 1995, PP. 296-317.
- [8] S. Y. Kung and K. I. Diamantraras, "Adaptive Principal Component Extraction (APEX) and applications", *IEEE Trans. on Signal Processing*, Vol. 42, N°5, May 1994, PP. 1202-1217.
- [9] P. Baldi and K. Hornik, "Neural networks and principal component analysis: learning from examples without local minima", *Neural Networks*, Vol. 2, 1989, PP. 53-58.
- [10] S. Bannour and M. Azimi-Sadjadi, "Principal component extraction using recursive least squares learning", *IEEE Trans. on Neural Networks*, Vol. 6, N°2, Mar. 1995, PP. 457-469.
- [11] J. Mao and A. K. Jain, "Regularization techniques in artificial neural networks", *Proc. of World congress on neural networks*, Portland, OR, July 1993, PP. IV75-IV79.

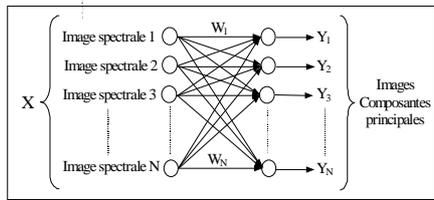


Fig. 1 : modèle du réseau de la première approche

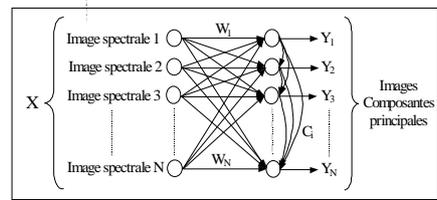


Fig. 2 : modèle du réseau de la deuxième approche

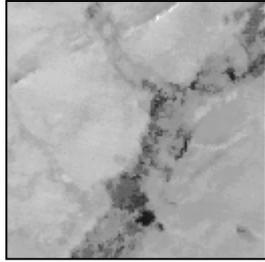


Image 1 : deuxième image spectrale du satellite

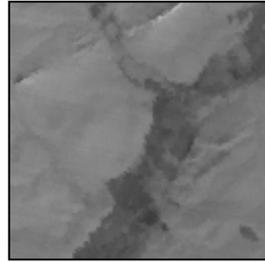


Image 2 : cinquième image spectrale du satellite

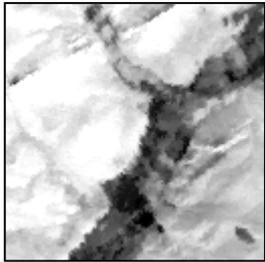


Image 3 : CP1 de la première approche

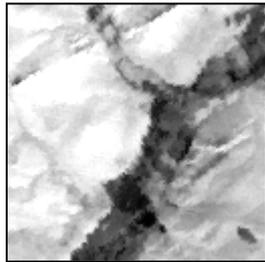


Image 4 : CP1 de la deuxième approche

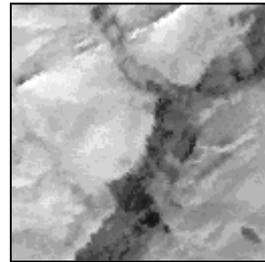


Image 5 : CP2 de la première approche

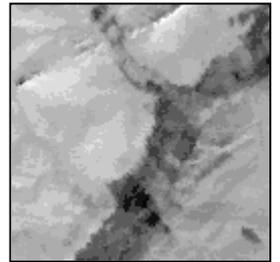


Image 6 : CP2 de la deuxième approche

Tab. 1 : le SNR dans les six images spectrales originales

Images	Image 1	Image 2	Image 3	Image 4	Image 5	Image 6
SNR	45.720	35.683	130.659	72.639	313.157	138.772

Tab. 2 : le SNR dans les premières et deuxièmes CP des deux approches neuronales

	Première approche		Deuxième approche	
Image composante principale (CP)	CP1	CP2	CP1	CP2
SNR	775.422	11.172	771.210	5.789

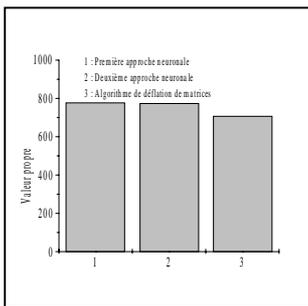


Fig. 3 : 1^{ère} valeur propre

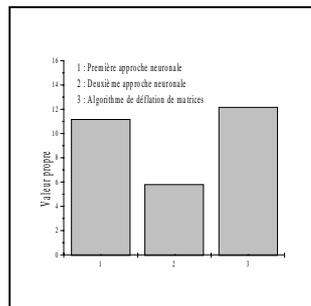


Fig. 4 : 2^{ème} valeur propre

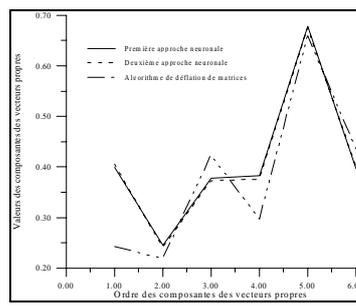


Fig. 5 : 1^{ère} vecteur propre

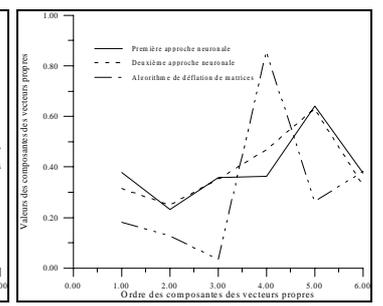


Fig. 6 : 2^{ème} vecteur propre

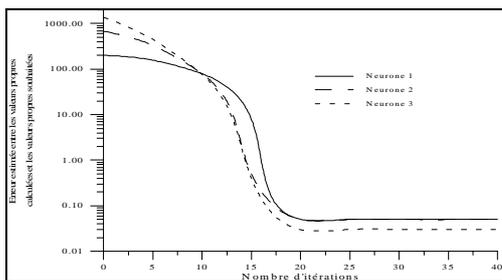


Fig. 7 : convergence de la première approche neuronale

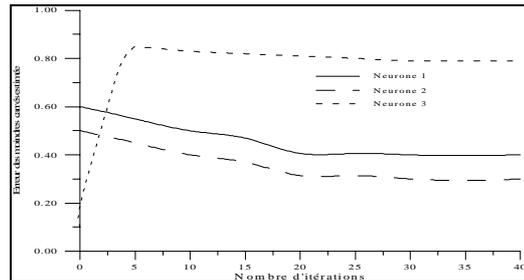


Fig. 8 : convergence de la deuxième approche neuronale