Télédétection, 2003, vol. 3, n° 2-3-4, p. 349–360
Droit de photocopie soumis à l'autorisation

© 2003 CONTEMPORARY PUBLISHING INTERNATIONAL
Publié sous l'enseigne *Éditions scientifiques GB*
Imprimé en France

FUSION CONTEXTUELLE D'IMAGES RSO DE ERS-1 ET DE TM DE LANDSAT-5

MOSTÉFA BELHADJ-AISSA, AICHOUCHE BELHADJ-AISSA* et RADJA KHEDAM.

Laboratoire de traitement d'images et rayonnements
Faculté de génie électrique
Université des sciences et de la technologie Houari Boumediene
Boîte Postale 32, Bab Ezzouar, 16111, Alger, Algérie.
Téléphone : 213 21 24 28 02 ; télécopieur : 213 21 24 28 02 ou 213 21 24 71 87
Courriel : h.belhadj@mailcity.com

(soumis le 22 mars 2002 ; révisé le 30 juillet 2002 ; accepté le 7 novembre 2002)

Les capteurs optiques et radar des satellites d'observation de la Terre fournissent une base de données caractérisées par différentes informations physiques mesurées à différentes échelles et à différentes périodes. Cette combinaison, à un niveau d'exploitation des images multisources, a permis aux utilisateurs de cette information d'établir des méthodologies qui prennent en charge les différents aspects de fusion de données et leur intégration dans les systèmes d'information géographique. Le travail que nous présentons porte sur la fusion contextuelle, au niveau décisionnel, des images RSO de ERS-1 et des images TM de Landsat-5 sur la région est d'Alger. La méthodologie proposée est la combinaison de l'information texturale de l'image radar, modélisée par la signature de texture déduite des matrices de cooccurrence des plans bits et de l'information multibande du même voisinage. Elle est décomposée en trois étapes essentielles. La première étape consiste en la détermination des signatures texturale et spectrale des échantillons des classes. La deuxième étape est la modélisation de l'information spectrale et texturale par les paramètres statistiques d'ordre 1 et 2 des échantillons. Afin de pondérer l'apport de chaque type d'information, nous avons introduit des facteurs de fusion dans le modèle décisionnel qui représentera la troisième partie. Une étape d'évaluation a été faite par les matrices de confusion sur des échantillons test. Pour cela, nous avons utilisé des photographies aériennes à l'échelle du 1/40 000 numérisées comme support pour la localisation des échantillons et l'évaluation qualitative des résultats.

Mots-clefs : fusion ; classification contextuelle ; matrice de cooccurrence des plans bit ; niveau décisionnel.

CONTEXTUAL FUSION OF IMAGES ROS (ERS-1) AND TM (LANDSAT-5)

The optical sensors and radar of the observation satellites of the Earth provide a data base characterized by various physical information measured at various scales and various periods. The combination, at the level of exploitation of the multi-source images made it possible to the users of this information to establish methodologies which deal with the various aspects of data fusion and their integration in the geographical information systems. Our work concern contextual fusion, at the decisional level, of ROS (ERS-1) and TM of Landsat-5 images on the area of Alger. The methodology suggested here is the combination of the textural information of the radar image, modelled by the signature of texture deduced from the matrices of co-occurrence of the plans bits and the multispectral information of the same vicinity. There are three essential stages : the first stage is the determination of the textural and spectral signatures of the samples of classes. The second stage is the modeling of spectral and textural information by the statistical parameters of order 1 and 2 of the samples. In order to balance the contribution of each type of information we introduce fusion factors into the decisional model which will represent the third part. An evaluation was made by the confusion matrix on test samples, air photographs at scale 1/40 000 were used as support for the localization of the samples and the qualitative validation of the results.

Keywords : fusion ; contextual classification ; co-occurence matrix of the bit plans ; decisional level.

1. INTRODUCTION

Les méthodes de classification dirigées ou non dirigées, basées seulement sur l'information radiométrique du pixel n'offrent pas toujours des résultats précis sur la représentativité des thèmes par les classes. L'imprécision de ce type de résultats est plus importante dans le traitement des images à haute résolution spatiale. Néanmoins, cette résolution fait augmenter la variabilité spatiale des points-images, ce qui se traduit par une forte corrélation interpixels et elle réduit la séparabilité statistique des classes (Byeungwoo and Landgrebe, 1992). Ces phénomènes conduisent à une dégradation des performances des méthodes de classification conventionnelles ou ponctuelles. Afin d'exploiter l'information apportée par les différentes sources, des approches de classification contextuelle et de fusion à plusieurs niveaux de traitement ont été proposées et testées par différents groupes de recherche (Byeungwoo and Landgrebe, 1999**)**. L'utilisation conjointe de données hétérogènes, issues de différents capteurs électromagnétiques (visible,

* Correspondant principal

proche infrarouge, micro-onde) requiert l'application de méthodes dites de fusion de données. Ce terme, quoique récent, désigne une démarche adoptée spontanément dans de nombreux domaines. Ces méthodes sont basées sur l'utilisation des principaux attributs qui caractérisent le pixel-image à savoir les attributs spatiaux (liés à la résolution spatiale), les attributs spectraux (liés au nombre d'images multibandes non corrélées issues de plusieurs capteurs) et les attributs temporels qui dépendent des images multidates (Wald, 2000). La fusion de données constitue un cadre formel dans lequel s'expriment les moyens et les techniques permettant l'alliance des données provenant de sources diverses. Elle vise à l'obtention d'informations de plus grande qualité; la définition exacte de « plus grande qualité » dépendra de l'application. Le travail que nous présentons a porté sur la fusion contextuelle, au niveau décisionnel, des images RSO de ERS-1 et TM de Landsat-5. La méthodologie proposée est la combinaison de l'information texturale (du contexte) de l'image radar modélisée par la signature de texture des matrices de cooccurrences des plans bits et de l'information multibande modélisée par les paramètres statistiques déduits des signatures spectrales du même voisinage. Le modèle de fusion de données a été établi à partir des données des échantillons identifiés et localisées sur des photographies aériennes numérisées. Le travail est structuré en trois parties. Nous présentons d'abord le problème de la fusion dans lequel nous précisons le niveau de fusion que nous avons considéré dans ce travail. Nous décrivons ensuite la région d'étude, ses caractéristiques et ses thèmes prédominants ainsi que le processus d'établissement de la base d'entraînement. Dans la deuxième partie nous présentons l'approche de fusion développée avec les niveaux d'extraction de l'information texturale et spectrale; des résultats partiels sont présentés et une interprétation visuelle est faite. La troisième partie est consacrée à l'évaluation quantitative à partir des matrices de confusion et aux paramètres statistiques qui en découlent.

1.1. Objectifs

Notre travail est une partie du projet « intégration des données de télédétection dans un système d'information géographique » pour la structuration optimale de l'espace urbain et la gestion des terres agricoles de la ville d'Alger. Le but visé est d'exploiter les résultats de la classification contextuelle avec des évaluations qualitative et quantitative. En effet, les images classifiées sont utilisées comme support pour la vectorisation des thèmes au sol. L'image classifiée (figure 1) représente toute la zone utilisée pour l'étude de la ville d'Alger avec la superposition de quelques couches thématiques. Nous avons appliqué la méthodologie développée sur une zone pilote contenant les principaux thèmes ciblés.

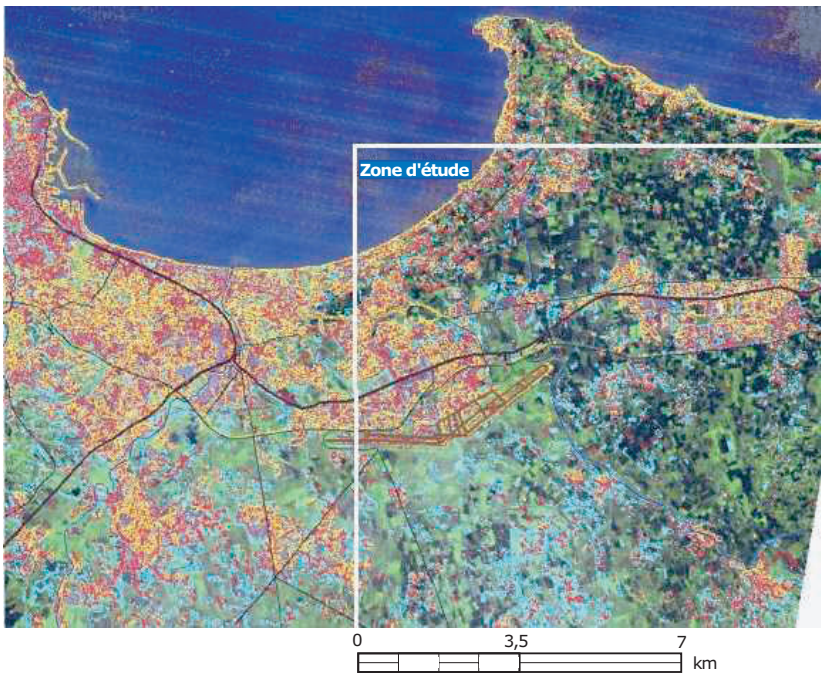


FIGURE 1 La baie d'Alger. *The bay of Alger*.

1.2. Contexte de la fusion de données

La ***classification contextuelle*** (ce terme associe souvent le contexte spatial ou temporel d'un pixel) peut avoir la même ***signification informationnelle*** que la ***fusion d'images***. Dans ce cadre, il s'agit de la fusion au niveau du modèle décisionnel par l'intégration des caractéristiques texturales. Le concept de la fusion de données a été abordé par plusieurs approches de traitement d'images. Si l'on se réfère aux trois niveaux dans lesquels s'opère la fusion (Pohl, 1997 ; Pohl and Van Genderen, 1998), qui sont le niveau du pixel, le niveau des attributs et le niveau de la décision, on dirait que la plupart des méthodes de fusion d'images multibandes, multirésolutions ou multidates traitent la réponse d'un pixel dans cet ensemble images en fonction d'une topologie donnée de son voisinage immédiat. Si le résultat de cette fusion est une partition ou un groupement en thèmes, il s'agit alors de la fusion au niveau décisionnel qui se ramène à une classification contextuelle.

2. RÉGION D'ÉTUDE

Les images multisources utilisées dans cette application sont issues des capteurs TM du satellite LANDSAT-5 et RSO du satellite ERS-1 acquises en mai 1996. Ces images couvrent la région est d'Alger, d'environ 16 km × 16 km. Les images sont géocodées et de résolution spatiale de 30 m × 30 m et l'image radar est ré-échantillonnée à la même résolution.

La région d'étude est située à environ 10 km à l'est d'Alger. Région initialement agricole, elle est aujourd'hui envahie par un tissu urbain qui se densifie de jour en jour. Elle est marquée par une forte concentration urbaine (figure 2) (partie centrale) autour de l'Université de Bab-Ezzouar et la commune de Dar-El-Beida ; entre les deux, on trouve des terrains nus. La partie gauche inférieure de l'image correspond à un tissu urbain moins dense à proximité de l'aéroport international et à une partie de la côte et la mer. Dans les différents tissus urbains, l'espace agricole s'est rétréci et on trouve tantôt des terrains cultivés et tantôt des sols nus, donc très peu de végétation. La description thématique, appuyée par les photographies aériennes à l'échelle du 1/40 000 acquises en 1999, permet d'identifier et de localiser les thèmes prédominants suivants : l'eau **(E)**, les zones urbaines denses **(UD)**, les zones urbaines moins denses **(UMD)**, les routes et les pistes d'aéroport **(SL)**, les régions agricoles **(RA)** et les sols nus **(SN)**. Cependant, lors du processus de classification, d'autres classes thématiques peuvent être prises en considération pour affiner les résultats obtenus.

3. EXTRACTION DE LA BASE D'ENTRAÎNEMENT

Sur l'image, on définit des parcelles (appelées aussi échantillons) d'apprentissage qui sont des zones de l'image que l'on juge représentatives des différents thèmes. Cette identification est effectuée par rapport à des données de référence qui peuvent être des cartes thématiques, des photographies aériennes numérisées et géocodées, des images déjà classifiées ou des données prises sur le terrain. Le choix des zones d'apprentissage mérite beaucoup d'attention. Deux facteurs essentiels doivent être pris en considération :

– la répartition des zones d'apprentissage d'une classe sur toute l'image à traiter; si deux régions distinctes spatialement représentent la même classe, il faudrait extraire une parcelle (ou plusieurs) de chaque région; ceci traduit au mieux la variabilité de la classe ;

– un nombre d'échantillons d'apprentissage appartenant à une classe est utilisé ; ce nombre doit être supérieur ou égal à $10(1+n)$, où n est le nombre de bandes spectrales de l'image à classifier.

Les zones d'apprentissage se présentent sous la forme de rectangles dont les pixels intérieurs sont ensuite extraits. On s'attachera tout au moins à isoler, à l'intérieur du contour, des pixels de valeurs les plus homogènes possibles en éliminant les pixels de bordure. Quant à l'apprentissage, il consiste, pour chaque classe donnée, à rassembler et à caractériser l'ensemble des pixels situés dans les zones d'apprentissage sélectionnées. Cette caractérisation nécessite une hypothèse sur la loi de distribution des pixels d'apprentissage et la taille des échantillons ou sous-échantillons qui doit être supérieure ou égale au voisinage (dimension du contexte spatial) considéré.

4. CLASSIFICATION CONTEXTUELLE

La classification d'un pixel de l'image est réalisée par la recherche d'une mesure de ressemblance entre ce pixel, dans son contexte spatial ou temporel, et les différents paramètres caractéristiques calculés durant la phase d'apprentissage pour chaque classe. Cette mesure de ressemblance est traduite d'un point de vue statistique par le calcul de la probabilité d'appartenance du pixel à chaque classe. Le processus de fusion contextuelle développé (Belhadj-Aissa, 1998) détermine le modèle de ressemblance à partir de deux paramètres qui sont la **signature de texture** et la **signature spectrale**. La signature texturale est déduite des matrices de cooccurrence des plans bits pour le voisinage du pixel à classer. Quant à la signature spectrale, elle est donnée par la réponse du pixel dans chaque canal spectral. Pour cela, nous présentons, dans une première étape, le principe d'extraction de la signature de texture.

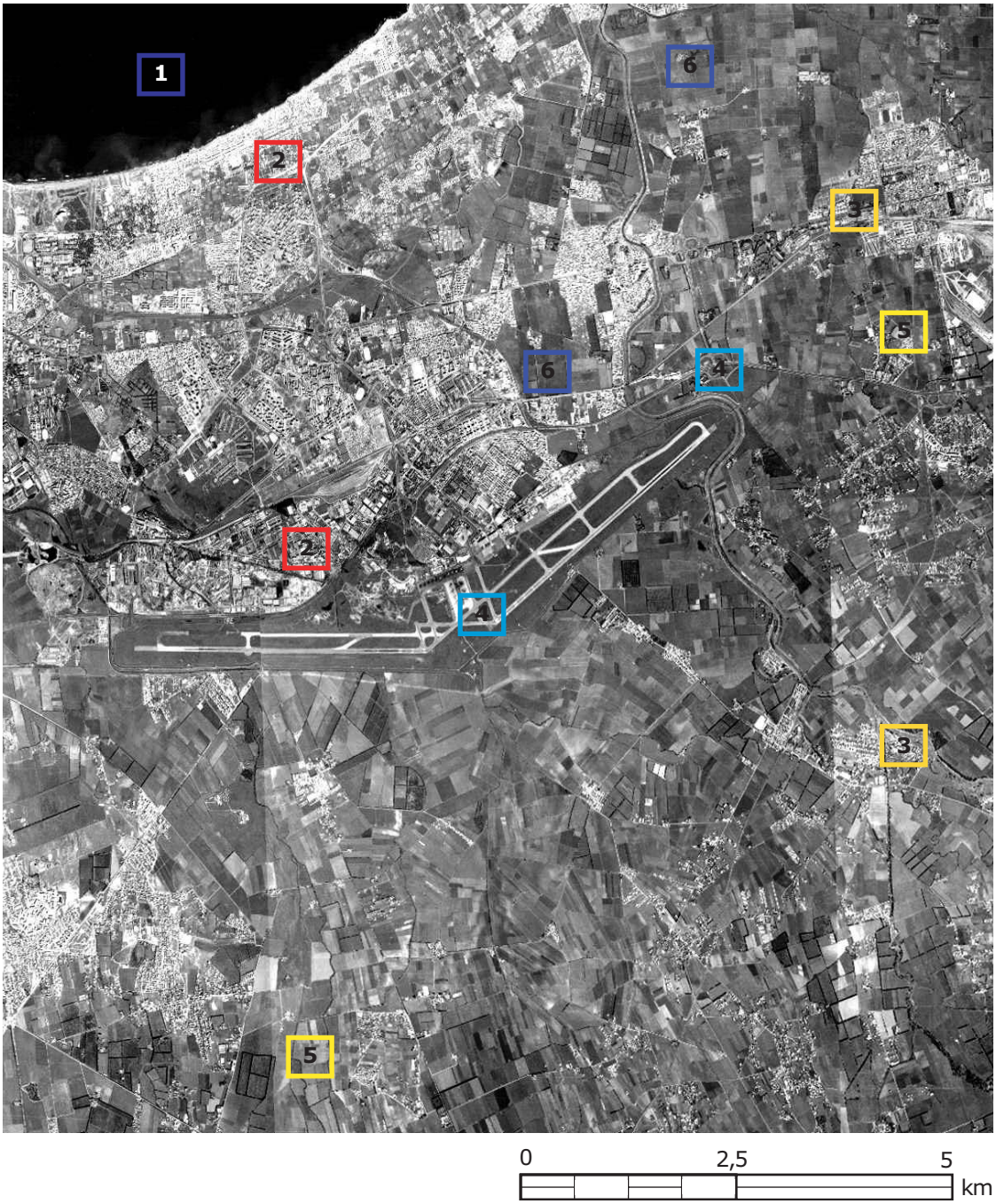


FIGURE 2 Localisation interactive des échantillons sur une image photographie aérienne. *Interactive localization of the samples on air photographs image.*

4.1. Contexte spatial et signature de texture

4.1.1. Contexte spatial

Soit le voisinage $V_{xy} = M \times M$ défini par l'ensemble des $\{ I_i(x,y)\}$ où I_i est la radiométrie du point de coordonnées (x,y) et i = 1, 2, ... $M \times M$. Chaque élément de $\{ I_i(x,y)\}$ est décomposé en : $I_i^p(x,y)$ où p représente le nombre des plans bits (p = 0, 1, 2…7).

La fonction $I_i^p(x,y)$ est binaire telle que

$$I_i^p(x,y) = \begin{cases} 1 \\ \text{où} \\ 0 \end{cases}$$

Le voisinage $V_{x,y}$ devient alors $\{I_i^p(x,y)\}$

Pour chaque plan bits, nous définissons les matrices de dépendance des points spatialement voisins où les éléments de celles-ci sont les fréquences d'occurrence des transitions (0,0), (0,1), (1, 0) et (1,1) suivant une relation R (d,θ), indiquant la distance et la direction de transition d'un point à un autre. Avec cette structure, les dimensions des matrices sont de (2×2) et les fréquences d'occurrence sont données par l'expression suivante :

$$F_V^p(i,j/R) = \text{card}\left\{ \begin{array}{l} (x_1,y_1),(x_2,y_2) \\ \in V_{x,y} / \; I^p(x_1,y_1) = i, \\ I^p(x_2,y_2) = j \quad \text{selon R} \\ \text{où} \;\; (i,j) = (0,0),(0,1),(1,0) \;\; \text{ou} \;\; (1,1) \end{array} \right\} \qquad (1)$$

Avec l'hypothèse de symétrie des matrices, le nombre d'éléments de chaque matrice se réduit à trois éléments : $\{f_v^p(0,0), f_v^p(0,1), f_v^p(1,1),\}$

4.1.2. Interprétation des fréquences des plans bits

Sur la figure 3, on représente les fréquences de transitions dans les plans bits calculés sur un voisinage de 3×3 extrait de l'image radar. Dans le cas général, les informations qualitatives obtenues de cet histogramme sont :

– pour un voisinage V_{xy} homogène, le graphe ne contient que les transitions (0,0) et/ou (1,1) ;

soit un voisinage homogène $V_{xy} = \begin{matrix} 91 & 91 & 91 \\ 91 & 91 & 91 \\ 91 & 91 & 91 \end{matrix}$;

les plans bits correspondants sont :

$$I_i^0 = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} ; \; I_i^1 = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} ;$$

$$I_i^2 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} ; \; I_i^3 = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} ;$$

$$I_i^4 = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} ; \; I_i^5 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} ;$$

$$I_i^6 = \begin{matrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{matrix} ; \; I_i^7 = \begin{matrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} ;$$

les fréquences d'occurrence de ce voisinage sont les suivantes :

$F^0(0,0) = 0 \quad F^1(0,0) = 0 \quad F^2(0,0) = 12 \quad F^3(0,0) = 0$
$F^4(0,0) = 0 \quad F^5(0,0) = 12 \quad F^6(0,0) = 0 \quad F^7(0,0) = 12$

$F^0(0,1) = 0 \quad F^1(0,1) = 0 \quad F^2(0,1) = 0 \quad F^3(0,1) = 0$
$F^4(0,1) = 0 \quad F^5(0,1) = 0 \quad F^6(0,1) = 0 \quad F^7(0,1) = 0$

$F^0(1,1) = 12 \quad F^1(1,1) = 12 \quad F^2(1,1) = 0 \quad F^3(1,1) = 12$
$F^4(1,1) = 12 \quad F^5(1,1) = 0 \quad F^6(1,1) = 12 \quad F^7(1,1) = 0$

on remarque que les fréquences des transitions (0,1) qui expriment la variabilité spatiale sont nulles :

– pour un voisinage fortement texturé, un ou plusieurs plans présentent de fortes fréquences au niveau des transitions (0,1) ;

– la dynamique du voisinage se traduit par la forme décroissante ou croissante des fréquences des transitions.

Les images « plans bits » sont des masques de thèmes différents. Elles peuvent aider à la localisation des classes. Sur les figures 4 et 5, nous présentons les images des plans bits 5 et 6 respectivement. Nous remarquons que le plan 6 met en évidence les autoroutes et le trait de côte, quant à l'image du plan 5, elle fait ressortir les zones urbaines.

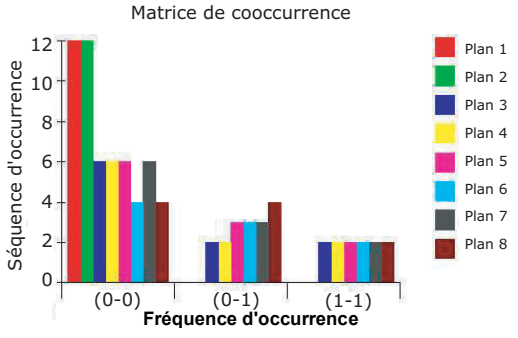


FIGURE 3 Fréquences d'occurrence des plans bits pour un domaine D. *Occurrence frequencies of the bits plans for a field D.*

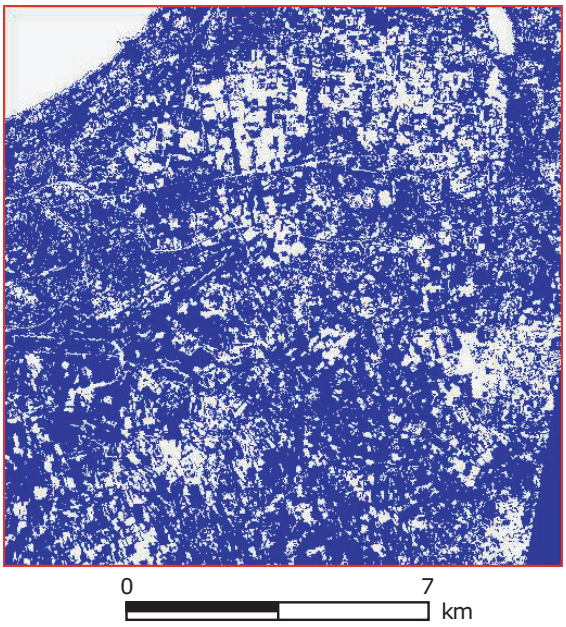


FIGURE 4 Image du plan bit 5. *Plan bit 5 image.*

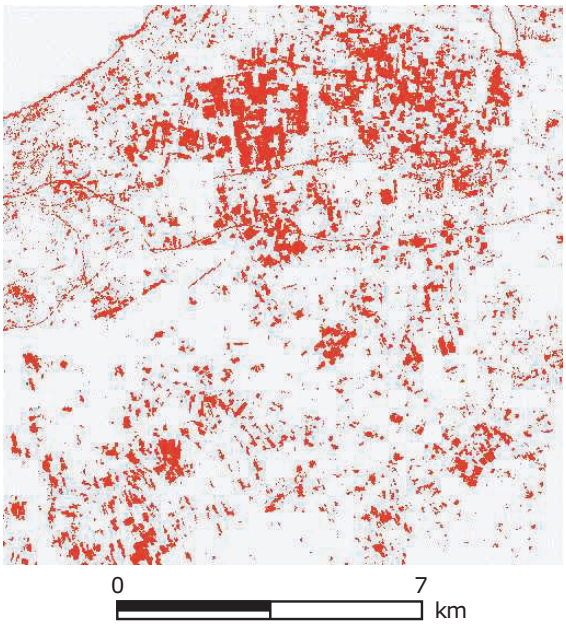


FIGURE 5 Image du plan bit 6. *Plan bit 6 image.*

4.1.3. Signature de texture

Si l'on a les éléments de chaque matrice de cooccurrence d'un plan bit donné $\{F_v^p(i,j/R)\}$, nous définissons la signature de texture du voisinage V_{xy} centré sur le point de coordonnées (x,y) par :

$$S_T(V) = \left[(F^0(i,j/R)(F^1(i,j/R)(F^2(i,j/R)...(F^7(i,j/R)\right]^T \quad (2)$$

$$S_T(V) = \begin{bmatrix} f^0(0,0) & f^1(0,0) & . & . & . & . & . & f^7(0,0) \\ f^0(0,1) & . & . & f^3(0,1) & . & . & . & . \\ f^0(1,1) & . & . & . & . & f^5(1,1) & . & f^7(1,1) \end{bmatrix}^T \quad (3)$$

Ce paramètre est défini sur une fenêtre et pour un canal donné. Or, dans la classification contextuelle, le nombre de variables n doit être supérieur ou égal à 2, ce qui conduit à la détermination de la signature texturale de la forme suivante :

$$S_T(v) = \left[S_T^1, S_T^2, ...S_T^n\right].$$

La dépendance spatiale est quantifiée par $S_T(V)$. Il reste à voir comment incorporer cette information pour établir le modèle de classification contextuelle. La classification étant dirigée, chaque échantillon d'une classe sera caractérisé par les paramètres statistiques d'ordre 1 et 2. À chaque position de la fenêtre d'analyse (taille de voisinage), une signature de texture est déterminée. La moyenne des signatures de texture sur tout l'échantillon, avec une fenêtre d'analyse, est donnée par :

$$\text{moy}(S_T/C_i) = \begin{bmatrix} \frac{1}{n_w}\sum_{i=1}^{n_w} f_i^0(0,0) & . & . & . & . & . & . \\ . & . & . & . & . & . & \frac{1}{n_w}\sum_{i=1}^{n_w} f_i^7(0,1) \\ \frac{1}{n_w}\sum_{i=1}^{n_w} f_i^0(1,1) & . & . & . & . & . & \frac{1}{n_w}\sum_{i=1}^{n_w} f_i^7(1,1) \end{bmatrix}^T \quad (4)$$

$$\text{moy}(S_T/C_i) = \begin{bmatrix} \mu_1^0 & . & . & \mu_1^3 & . & . & . \\ . & . & . & . & . & . & \mu_2^7 \\ \mu_3^0 & . & . & . & . & . & \mu_3^7 \end{bmatrix}^T \quad (5)$$

où n_w est le nombre de fenêtres par échantillon, $n_w = (K-(M-1))\cdot(L-(M-1))$ où $K \times L$ est la taille de l'échantillon de la classe et $M \times M$ la taille de la fenêtre d'analyse. Le deuxième paramètre est représenté par l'écart type des signatures de texture pour un échantillon de la classe Ci :

$$\text{sd}(S_T/C_i) = \begin{bmatrix} \left(\frac{1}{n_w}\sum_{i=1}^{n_w}\left(f_i^0(0,0)-\mu_1^0\right)^2\right)^{1/2} & . & . & . & . & . & . \\ . & . & . & . & \left(\frac{1}{n_w}\sum_{i=1}^{n_w}\left(f_i^6(0,1)-\mu_2^6\right)^2\right)^{1/2} & . & . \\ . & . & . & . & . & . & \left(\frac{1}{n_w}\sum_{i=1}^{n_w}\left(f_i^7(1,1)-\mu_3^7\right)^2\right)^{1/2} \end{bmatrix}^T \quad (6)$$

$$sd(S_T / C_i) = \begin{bmatrix} \sigma_1^0 & \cdots & \cdots \\ \cdots & \cdots & \sigma_2^6 \\ \cdots & \cdots & \sigma_3^7 \end{bmatrix}^T \quad (7)$$

Les hypothèses lors de la mise en œuvre sont les suivantes :

– $I_i(x,y)$ est une variable aléatoire qui suit une distribution normale multivariable ;

– les échantillons forment la base de données pour l'estimation des paramètres du modèle de décision ;

– dans la règle de classification, les fonctions discriminantes doivent être déterminées en fonction des signatures spectrale/spatiale.

Les contraintes de la méthode contextuelle dirigée sont les suivantes :

– un échantillon $E(C_i)$ de la classe C_i doit avoir une structure carrée ou rectangulaire pour constituer une topologie de voisinage et doit conserver son organisation spatiale ;

– un échantillon peut être formé de plusieurs sous-sites $\mathbf{e_m(Ci)}$, c'est-à-dire que $E(Ci) = \{e_m(Ci)\}$ pour assurer la représentativité de la classe C_i (répartition de l'échantillon sur toute l'image) ;

– la taille des sous-sites $\mathbf{e_m(Ci)}$, qui est parfois égale à E(Ci), doit être supérieure ou égale à la fenêtre d'analyse (dimension du voisinage).

4.2. Contexte spectral

Nous définissons les mêmes paramètres statistiques, la moyenne et l'écart type, sur la fonction radiométrique des l'échantillons de chaque classe. Soit l'ensemble des fonctions *{Ii(E/Ci)}* de l'échantillon *E* qui correspond à la classe C_i : le vecteur moyenne sur l'espace des variables images est :

$$moy(E/Ci) = \underbrace{\begin{bmatrix} \frac{1}{K \times L} \sum_{j=0}^{Ng-1} I_j^1(E/C_i) & & \\ & \frac{1}{K \times L} \sum_{j=0}^{Ng-1} I_j^2(E/C_i) & \\ & & \cdots \frac{1}{K \times L} \sum_{j=0}^{Ng-1} I_j^n(E/C_i) \end{bmatrix}^T}_{= \left[\mu_i^1, \mu_i^2, \ldots\ldots\ldots \mu_i^n\right]^T} \quad (8)$$

L'écart type est :

$$s_d(E/Ci) = \underbrace{\begin{bmatrix} \left(\frac{1}{K \times L} \sum_{j=0}^{Ng-1} I_j^1 \left((E/C_i) - \mu_i^1 \right)^2 \right)^{1/2} & & \\ & \cdots\cdots \left(\frac{1}{K \times L} \sum_{j=0}^{Ng-1} I_j^n \left((E/C_i) - \mu_i^n \right)^2 \right)^{1/2} & \end{bmatrix}^T}_{= \left[\sigma_i^1, \sigma_i^2, \ldots\ldots\ldots\ldots \sigma_i^n\right]^T} \quad (9)$$

où n est le nombre de canaux à fusionner.

4.3. Modèle de décision

L'étape finale de la méthode est l'établissement du modèle décisionnel de classement. Pour cela, nous nous sommes basés sur une mesure de dispersion *(pixel-classe)* exprimée sous forme de distance statistique. Le modèle est formulé comme suit : soit le pixel de coordonnées *(x,y)*, de voisinage V_{xy} et de fonction radiométrique *I(x,y)* ; nous cherchons à classer ce pixel dans une classe qui lui ressemble le plus au sens de *{Vxy, I(x,y)}* en déterminant la distance minimale donnée par :

$$d\left(I(x,y) \in V_{xy}, C_i\right) = a \cdot G_T + b \cdot G_I \quad (10)$$

avec :

$$G_T(C_i) = \sum_n \left(\sum_{j=0}^{p} \left(\sum_{k=1}^{3} \frac{\left| f_k^j(V_{xy}) - \mu_k^j \right|}{\sigma_k^j} \right) \right) \text{ et}$$

$$G_I(C_i) = \sum_n \left(\frac{\left| I^n(x,y) - \mu_i^n \right|}{\sigma_i^n} \right)$$

f_k^j exprime les fréquences d'occurrence des transitions : k =1 pour (0,0), k = 2 pour (0,1) et k = 3 pour (1,1), j = 0, p = 7.

Les constantes a et b sont des facteurs de pondération des effets texturaux et spectraux ; elles sont définies par l'utilisateur en fonction du type d'image.

La fonction discriminante associée à cette distance est définie par :

$$g_{Ci}\left(I(x,y) \in V_{x,y}\right) = -d(I(x,y) \in V_{xy}, C_i) \quad (11)$$

La règle de décision est la suivante :

le point $(I(x,y) \in V_{x,y}) \in C_i$

$$\text{Si } g_{Ci}(I(x,y) \in V_{x,y}) > g_{Cj}(I(x,y) \in V_{x,y}) \quad (12)$$

pour tout $i \neq j$

4.4. Application

La classification contextuelle a été appliquée par fusion des images optiques pour l'aspect spectral et de l'image radar pour l'aspect textural. L'extraction de l'information de texture peut être faite, soit à partir d'un canal spectral approprié, soit du premier axe d'une analyse en composantes principales appliquée sur les données multibandes ou bien à partir d'une image radar. L'utilisation de l'image radar présente certains avantages à savoir :

– la signature texturale représente l'information de rugosité (structure spatiale) d'un thème ;

– l'optimisation (en complexité d'algorithme et en temps de calcul) du processus de détermination des paramètres de texture (avec n canaux, la procédure est appliquée n fois).

Quant aux inconvénients de l'utilisation de l'image radar, ils relèvent des problèmes de la fusion de

données qui sont la superposition géométrique et la calibration radiométrique (au sens date de prise de vue). Le processus développé a été testé sur les images couvrant la région d'Alger. La description thématique de cet espace est plus complexe et il présente, dans la partie ouest, un terrain très accidenté. Pour cette partie du travail, nous présentons les images des résultats sur la zone d'étude décrite ci-dessus, de taille 512 × 512 pixels, avec des fenêtres d'analyse de tailles de 3 × 3 et 5 × 5 ainsi que des paramètres de pondération différents.

4.5. Discussion

Les résultats de la classification contextuelle dépendent des paramètres essentiels suivants : la taille du voisinage d'analyse, la représentativité des échantillons terrain et la nature des données à fusionner (normalisations radiométrique et géométrique). Si la fenêtre d'analyse est trop petite, l'information spatiale peut être insuffisante pour l'identification des thèmes différemment texturés et, dans le cas d'une fenêtre plus large, elle peut provoquer une confusion de thèmes. Alors, le choix de la taille est fait sur la base d'une connaissance a priori de la répartition des thèmes. Dans certaines applications, la taille de la fenêtre dépendra des classes que nous voulons obtenir avec une meilleure précision. D'après les résultats obtenus, nous constatons que la classification présente un aspect « sel et poivre » pour une fenêtre de 3 × 3 (figure 6) et des zones beaucoup plus homogènes pour une fenêtre de 5 × 5 (figure 9). Quant à la représentativité des échantillons, une meilleure connaissance de la réalité de terrain donnerait une classification optimale. Pour les paramètres de pondération spatiale-spectrale, nous remarquons que l'apport de la texture dans la classification contextuelle est mis en évidence par la présence des structures de certains thèmes, comme nous le constatons au niveau de la zone encadrée sur les figures 7 et 8. Par ailleurs, la précision des résultats de ce processus de fusion dépend de la direction d'analyse (0°,45°,90°,135°) pour l'obtention de la signature de texture. Une étude adéquate de la corrélation entre ces facteurs améliorerait la précision des résultats.

5. ÉVALUATION QUANTITATIVE

L'évaluation de la correspondance entre le résultat de la classification et la réalité de terrain est une étape importante pour l'exploitation de ces résultats dans les applications ultérieures. Elle se base sur des critères relatifs à la nature des données et à la méthode de classification utilisée, et s'applique sur les attributs des classes qui sont :

– la nature qualitative ou quantitative des classes ;

– la fréquence d'erreur d'affectation, pour une classe ou l'ensemble des classes ;

– l'importance des erreurs d'affectation, c'est-à-dire une erreur d'affectation dans une catégorie (urbain dense, urbain moins dense) ou entre catégories (classe urbaine, classe végétation) ;

– la qualité intrinsèque des données due aux conditions d'acquisition et à la résolution spatiale des capteurs satellitaires.

La précision de la classification est le résultat de cette évaluation. Une des méthodes d'évaluation la plus utilisée est la matrice de confusion ou matrice d'erreurs.

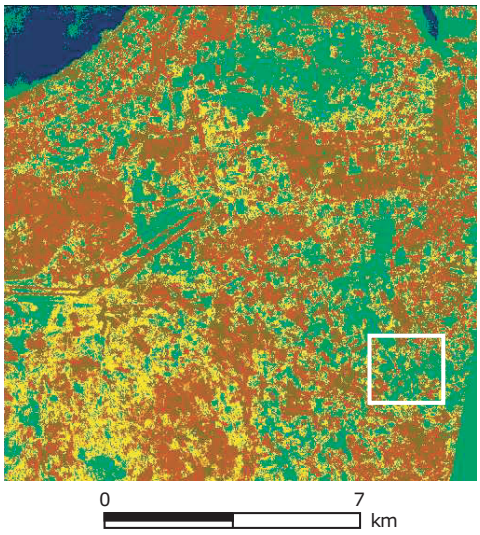


FIGURE 6 Image des résultats de la classification avec une fenêtre de 3 × 3, a = 0,4 , b = 1. *Image of the classification's results with a 3 × 3 window, a = 0,4 , b = 1.*

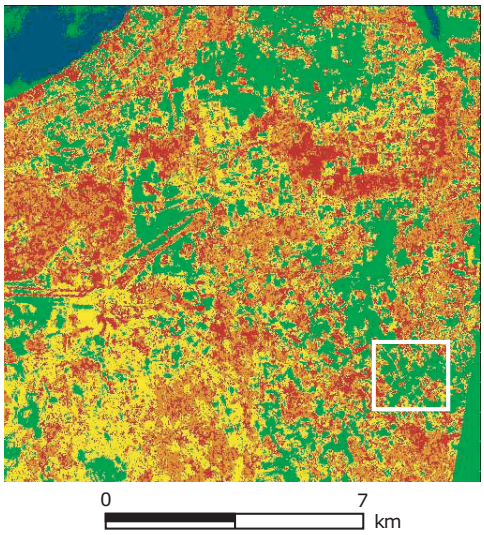


FIGURE 7 Image des résultats de la classification avec une fenêtre de 3 × 3, a = 0,6, b = 1. *Image of the classification's results with a 3 × 3 window, a =0,6, b = 1.*

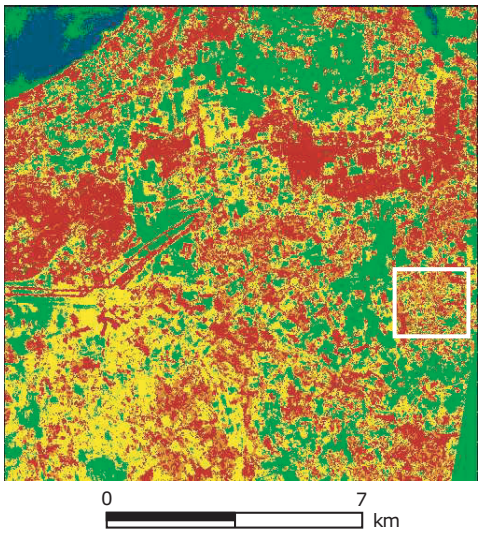


FIGURE 8 Image des résultats de la classification avec une fenêtre de 3×3 (0°) a = 0,8, b = 1. *Image of the classification's results with a 3 × 3 window, a = 0,8, b = 1.*

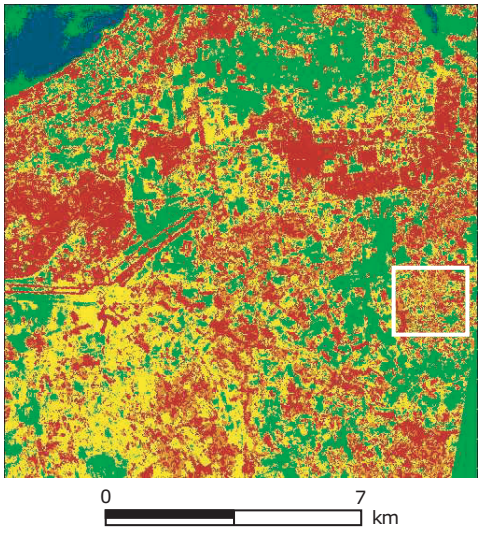


FIGURE 9 Image des résultats de la classification avec une fenêtre de 5×5 (0°) a = 0,8, b = 1 et une base d'entraînement différente (6 classes). *Image of the classification's results with a 5 × 5 window (0°) a = 0,8, b = 1 and different training basis (6 classes).*

5.1. Matrice de confusion

Si l'on dispose d'une réalité de terrain sur l'ensemble de l'image, la matrice de confusion constituerait un bon critère d'évaluation de la qualité d'une classification. C'est une matrice établie à partir de l'image classifiée et la réalité de terrain. Elle permet non seulement d'estimer le taux de bonne classification mais également les erreurs d'exclusion (ou d'omission) relatives au producteur de la classification et les erreurs d'inclusion (ou commission) relatives à l'utilisateur du résultat de la classification. La matrice de confusion est une matrice carrée de taille égale au nombre de classes et dont les éléments représentent le nombre de pixels affectés à chaque classe relativement à leur appartenance aux classes de la réalité de terrain. Dans la matrice, on dispose arbitrairement en colonnes les données de référence et en lignes le résultat de la classification. Une classification idéale produit une matrice diagonale (Congalton, 1991).

5.2. Indicateurs de précision

Soit X_{ij} les éléments de la matrice de confusion, X_{i+} la somme totale des éléments en lignes, X_{+i} la somme totale des éléments en colonnes, X_{ii} les éléments diagonaux, N le nombre total des pixels de la matrice, M le nombre de classes considérées. À partir d'une matrice de confusion, nous pouvons calculer quatre indicateurs de précision.

1) Précision-producteur :

$$PP = \frac{X_{ii}}{X_{+\mathrm{i}}} \quad (13)$$

Cet indicateur montre que les données de référence se retrouvent correctement dans leurs classes prédéfinies. Il donne la précision pour le producteur de la classification qui est intéressé de savoir comment il a classifié certains pixels relativement aux données de référence.

2) Précision utilisateur :

$$PU = \frac{X_{ii}}{X_{i+}} \quad (14)$$

Cet indicateur est une valeur calculée par la division du nombre des pixels correctement classés dans une classe par le nombre total des pixels qui se retrouvent dans cette classe.

3) Précision moyenne ou taux moyen de bonne classification :

$$\mathrm{PMC} = \frac{\sum_{i}^{M} X_{ii}}{N} \quad (15)$$

Cet indicateur est une mesure globale représentant le taux moyen de bonne classification. Notons que ce taux ne s'affranchit pas des différences de représentativité des classes dans les données de référence. Une classe influencera donc ce taux d'autant plus que sa représentativité dans les données de référence est importante.

4) Paramètre KAPPA ou KHAT

Le critère statistique KAPPA prend en compte la précision pour le producteur et la précision pour

l'utilisateur. Il est calculé à partir d'une matrice de confusion établie entre l'image classifiée et la réalité de terrain de toute l'image. Il est donné par l'expression suivante :

$$\hat{K} = \frac{N\sum_{i=1}^{M} X_{ii} - \sum_{i=1}^{M}\left(X_{i+} * X_{+i}\right)}{N^2 - \sum_{i=1}^{M}\left(X_{i+} * X_{+i}\right)} \quad (16)$$

Si la matrice de confusion, établie entre l'image classifiée et une réalité de terrain, ne représente que quelques parcelles homogènes de l'image, alors l'expression précédente est celle du critère statistique le plus souvent utilisé qui est le KHAT :

$$\hat{K}_i = \frac{NX_{ii} - \left(X_{i+} * X_{+i}\right)}{NX_{i+} - \left(X_{i+} * X_{+i}\right)} \quad (17)$$

Pour l'évaluation des résultats de la fusion ou la classification contextuelle, nous calculons le KHAT global et le KHAT de chaque classe. Les données de référence nécessaires pour le calcul de la matrice de confusion doivent être totalement ou partiellement différentes de celles ayant servi à l'élaboration de l'image classifiée qu'on veut évaluer (Girard *et al.*, 1999). Si l'on ne dispose pas de réalité de terrain, ce qui est généralement le cas, il faut procéder à une étape de définition d'une réalité de terrain qu'on utilisera ensuite pour établir la matrice de confusion. À cet effet, une partition de l'ensemble d'apprentissage W en deux sous-ensembles est effectuée pour constituer : un sous-ensemble de données d'apprentissage W_1 pour le classificateur et un sous-ensemble de données test W_2 à partir duquel sera constituée la matrice de confusion.

5.3. Application

Dans un fichier de contrôle, nous avons défini un ensemble d'échantillons pour chaque classe prédéfinie dans le processus de classification. Les classes sont :

Classe 1 : Eau **(E)**
Classe 2 : Urbain moins dense **(UMD)**
Classe 3 : Urbain dense **(UD)**
Classe 4 : Routes et pistes d'aéroport (structures linéaires **SL)**
Classe 5 : Régions agricoles **(RA)**
Classe 6 : Sol nu **(SN)**

Nous déterminons ensuite la matrice de confusion entre les échantillons étiquetés dans l'image classifiée en fonction des paramètres suivants : fenêtre d'analyse de 5 × 5, direction 0°, a = 0,8, b = 1, nombre d'échantillons par classe variable. Cette matrice est présentée sur le tableau 1. À partir de cette matrice, nous avons calculé les paramètres de précision suivants : le pourcentage de pixels correctement classifiés PMC = 72,8 % et le KHAT global = 62,5 % . Nous constatons que la valeur du coefficient KHAT est inférieure à celle du PMC. En effet, le coefficient KHAT est calculé en prenant en considération tous les éléments de la matrice, ce qui donne plus d'informations sur les erreurs d'affection pour chaque classe. Pour avoir une mesure de précision individuelle des différents thèmes, nous avons procédé au calcul des coefficients **P**récision de l'**U**tilisateur (**PU**), **P**récision du **P**roducteur (**PP**) ainsi que les erreurs d'omission et de commission pour chaque classe. Les résultats sont donnés sur le tableau 2. Nous remarquons que les erreurs de commission et d'omission sont nulles pour la classe Eau, quant à la classe Urbain moins dense, l'erreur de commission est égale à 81,9 %, ce qui signifie qu'il y a une grande partie des pixels appartenant à la classe 2 et qui sont affectés après classification dans d'autres classes. La précision de la classification contextuelle pour l'urbain dépend de la résolution spatiale des images, de l'hétérogénéité du pixel lui-même, de la taille du voisinage qui introduit un l'effet de bordure plus accentué pour les zones urbaines. Pour améliorer la précision d'une classification contextuelle, plusieurs tests doivent être effectués en modifiant la taille et la forme des échantillons (forme polygonale), la taille de la fenêtre d'analyse avec l'analyse des quatre connexités au lieu des huit connexités et, enfin, l'utilisation d'images à haute résolution (20 m, 10 m et 4 m).

5.3.1. Précision en fonction de la taille de la fenêtre

Nous avons déterminé le paramètre KHAT en fonction de la taille de fenêtre d'analyse pour une direction 0°et nous avons obtenu les résultats suivants : pour une fenêtre de 7 × 7, le KHAT est de 51 %, pour une fenêtre de 5 x 5, le KHAT est de 49 % et pour une fenêtre de 3 × 3 le KHAT est de 46 %. Nous pouvons en conclure que la précision de la classification contextuelle dépend fortement de la taille des primitives de la texture d'une classe, qui se traduit par la taille de la fenêtre d'analyse.

5.3.2. Précision en fonction de la direction d'analyse

Pour les mêmes sites d'entraînement avec une fenêtre d'analyse de 5 × 5, a = 0,8 et b = 1, nous avons calculé les paramètres PMC et le KHAT pour les quatre directions principales et nous avons obtenu les résultats présentés sur le tableau 3. D'après ces résultats, en fonction de la direction, les valeurs des paramètres sont sensiblement les mêmes. En effet, l'image test ne présente pas globalement de direction privilégiée : c'est un indicateur d'isotropie.

TABLEAU 1 Matrice de confusion pour 6 classes avec une fenêtre d'analyse de 5 × 5, direction 0°, a = 0,8, b = 1. *Confusion matrix for 6 classes with a 5 × 5 analysis window, direction 0°, a = 0,8, b = 1.*

Classes / Terrain	1	2	3	4	5	6	Total/ligne
1	**912**	0	0	0	0	0	912
2	0	**67**	3	13	4	14	101
3	0	0	**197**	0	20	1	218
4	0	109	11	**59**	15	35	229
5	0	2	51	22	**69**	14	158
6	0	192	2	16	2	**101**	313
Total/col.	912	370	264	110	110	165	**1931**

TABLEAU 2 Paramètres de mesure de la précision de chaque classe. *Measurement parameters of each class precision.*

	PU %	*PP %*	Erreur d'omission (100-PU)	Erreur de commission (100-PP)
Classe 1(E)	100	100	0	0
Classe 2 (UMD)	66,3	18,1	33,7	81,9
Classe 3 (UD)	90,4	74,6	9,6	25,4
Classe 4(SL)	25,8	53,6	74,2	46,4
Classe 5 (RA)	43,7	62,7	56,3	37,3
Classe 6 (SN)	32,3	61,2	67,7	38,8

TABLEAU 3 Variation du KHAT en fonction de la direction d'analyse. *Variation of the KHAT according to the direction of analysis.*

L'angle par rapport à l'horizontale ( )	0°	45°	90°	135°
KHAT (%)	48,3	47,4	48,5	47,8
PMC (%)	59	58,3	59,2	58,7

5.3.3. Précision en fonction des paramètres a et b

Nous avons effectué une étude sur la précision globale à l'aide du coefficient KHAT en fonction des paramètres de pondération. En pratique, nous avons fixé $b = 1$, nous avons fait varier le paramètre de pondération de la texture a et nous avons obtenu la courbe donnée sur la figure 10. Nous constatons que la précision exprimée par le KHAT augmente proportionnellement avec le paramètre a ($0 \leq a \leq 1$). Lorsque a est supérieur à un, la contribution de l'information texturale est plus importante que l'information spectrale, la classification converge vers une classification texturale (deux thèmes ayant une structure spatiale proche sont classés dans la même classe de texture). À cet effet, la précision du KHAT basée sur l'analyse contextuelle diminue.

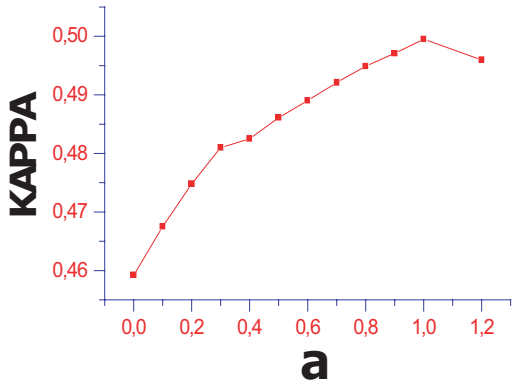


FIGURE 10 Variation de KAPPA en fonction de **a.** *Variation of KAPPA according to* ***a.***

5.3.4. Calcul adaptatif des paramètres a et b

Le concept de paramètres adaptés au type de données traitées est sans doute l'axe de recherche le plus complexe, mais aussi le plus intéressant dans le domaine de la classification des données satellitaires. Le paramètre **a** contrôle le degré de contribution de la texture dans la classification, dans une fenêtre donnée. Or, la texture est un phénomène non stationnaire et qui est variable spatialement. Il serait donc intéressant de le calculer dans cette fenêtre en fonction des éléments de cette dernière. Ainsi, nous proposons une méthode de calcul adaptatif du paramètre de texture **a** comme suit :

$$a_i = \frac{1}{8\,n\,(n-1)} \sum_{i=0}^{7} \mathrm{f}(0,1) \quad (18)$$

Dans la classification contextuelle, la contribution de l'information spectrale restera la même quelle que soit la nature de la texture (zone fortement ou peu texturée) et, pour cela, on prend b = 1. Cependant, pour une classification purement texturale, la règle de décision ne dépendra que du paramètre a, et de ce fait b = 0.

6. CONCLUSION

Vu que la classification contextuelle prend en compte la répartition des thèmes et les réponses spectrales dans les images traitées, les images résultantes présentent des classes disjointement hétérogènes. Cette hétérogénéité est due au fait qu'un thème, dont la taille est inférieure à la résolution spatiale du capteur, contribue avec un certain pourcentage à la formation d'un pixel-image. En outre, les informations contextuelles permettent d'extraire les indicateurs et les critères nécessaires pour l'extrapolation sur une superficie plus importante de la zone d'étude. La complémentarité avec les images thématiques peut aider le thématicien à établir une première description de l'état de surface, c'est-à-dire que la région est globalement constituée de parcelles de types agricole, urbain ou de sol nu. Dans ce cas, les classes sont définies en fonction de leurs propriétés radiométriques. Pour affiner la description des classes, on fait appel à la texture et à la structure de celle-ci, d'où la nécessité de fusionner ces deux informations pour toute application thématique utilisant, comme base de données, les images satellitaires optiques et radar. L'approche de fusion, présentée dans cet article introduit la notion de texture par les matrices de cooccurrence des plans bits. Parallèlement, nous avons abordé la notion contextuelle par les approches génétiques. Un processus de classification multisource basé sur les algorithmes est en cours d'évaluation avec une étude comparative des résultats obtenus par ces différentes approches.

Les images résultantes de la fusion sont utilisés comme un support quantitatif pour la réalisation de couvertures thématiques et leur intégration dans un système d'information géographique pour la gestion et la mise à jour de l'espace d'étude.

Remerciements

Les auteurs tiennent à remercier Alain Hillion de l'ENST de Bretagne, Yves Auda du Laboratoire d'écologie terrestre de Toulouse et Jean-Marie M. Dubois du CARTEL ainsi que tous les évaluateurs anonymes pour leur lecture critique du texte et leurs remarques.

Références bibliographiques

Belhadj-Aissa, A. (1998) Contribution au logiciel d'analyse et de traitement d'images satellitaires (L.A.T.I.S). Analyse spatiale-spectrale d'images satellitaires appliquées à la cartographie thématique. Thèse de doctorat d'état, Université des Sciences et de la Technologie Houari Boumediène, Bab Ezzouar, Alger, 268 p.

Byeungwoo, J. and Landgrebe, D. A. (1992) Classification with Spatio-temporel Interpixel Class Dependency Contexts. IEEE Transactions on Geoscience and Remote Sensing, vol. n°30, 4, p. 663-672.

Byeungwoo, J. and Landgrebe, D. A. (1999) Decision Fusion Approach for Multitemporal Classification. IEEE Transactions on Geoscience and Remote Sensing, vol. 37, n°3, p. 1227-1233.

Congalton, R. G., Oderwald, R. G. and Mead, R. A. (1983) Assessing Landsat Classification Accuracy Using Discrete Multivariate Analysis Statistical Techniques. Photogrammetric Engineering and Remote Sensing, vol. 49, n° 12, p. 1671-1678.

Congalton, R. G. (1991) A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, n°37, p. 35-46.

Girard, M.C. et Girard, C. (1999) Traitement des données de télédétection. Édition Dunod, Paris, 529 p.

Jhung, Y. and Swain, P. H. (1996) Bayesian Contextual Classification Based on Modified M-Estimates and Markov Random Fields. IEEE Transactions on Geoscience and Remote Sensing, vol. 34, n°1, p. 67-74.

Pohl, C. (1997) Operational issues of data fusion for visual image exploitation. Expert Meeting on Data Fusion Techniques, 8-9 décembre 1997, Freiburg, Germany, p. 10-15.

Pohl, C. and Van Genderen, J.L. (1998) Multisensor image fusion in remote sensing: concepts, methods and applications. International Journal of Remote Sensing, vol. 19, n° 5, p. 823-854.

Wald, L. (2000) A Conceptual Approach to the Fusion of Earth Observation Data. Survey in Geophysics, vol. 21, p. 177-186.