

CLASSIFICATION MULTISOURCES EN TÉLÉDÉTECTION PAR LA THÉORIE DE L'ÉVIDENCE AVEC L'INTÉGRATION DU CONTEXTE SPATIAL

H. Bracker, A. Hillion, M.-C. Mouchot

ENST de Bretagne
Dép. Images et Traitement d'Information
BP 832
29285 Brest Cédex

RÉSUMÉ

Le problème traité dans cet article s'inscrit dans le cadre général de la classification multisources en télédétection. Pour tenir compte de l'hétérogénéité des données et de l'incertitude liée à leur interprétation, nous avons choisi la théorie de Dempster/Shafer. Puisque les classes d'intérêt se présenteront sous la forme de régions homogènes dans la scène observée, le contexte spatial représente une source d'information supplémentaire. Pour exprimer cette contrainte de l'homogénéité spatiale, nous utilisons la modélisation markovienne. Elle nous fournit la distribution de probabilité jointe des classes. Nous l'exprimons dans le cadre évidentiel afin de la fusionner avec les observations fournies par les différentes images. Un nouveau critère de décision accélère la convergence de la classification itérative.

1 INTRODUCTION

Pour une meilleure discrimination des classes contenues dans une scène, nous utilisons des images multispectrales. Elles représentent la réflectivité du terrain observé dans différents domaines de longueur d'onde. Au préalable, une analyse statistique des images est nécessaire. Elle mène à la caractérisation des N classes d'intérêt θ_i par deux grandeurs: la densité a posteriori $p_m(x|\theta_i)$ - dépendant de la bande m considérée - et la probabilité d'apparition $p(\theta_i)$. Les deux grandeurs peuvent être déterminées de manière dirigée (à l'aide des connaissances a priori) ou non-dirigée (par des algorithmes pour la répartition des mélanges comme le EM, SEM). La classification sera effectuée selon la règle du Maximum de Vraisemblance (MdV). C'est-à-dire qu'un pixel est associé à la classe θ_k si

$$k = \max_i \left\{ p(\theta_i) \prod_{m=1}^M p_m(x|\theta_i) \right\}$$

où M est le nombre d'images à fusionner.

La théorie probabiliste est bien appropriée pour la description des jeux de données de nature homogène (statistiquement) et complète (spatialement). Toutefois, ces suppositions sur les données ne sont généralement pas remplies en télédétection. A cause des informations manquantes (nombre de classes dans le terrain, évolution de la signature spectrale des classes...), l'analyse des données sera déficiente et ne nous donnera qu'une vague idée de la réalité.

Puisque la théorie des probabilités de Bayes requiert une connaissance exhaustive (sur les classes contenues dans la scène observée) et précise (sur les lois de probabilités), une intégration des données déficientes polluera les informations plus précises et dégradera la performance du système de reconnaissance. Pour éviter cela et profiter quand même de ce type d'information, il est nécessaire de se placer dans un cadre mathématique plus général.

ABSTRACT

The subject of our article is the general problem of classification of multisource data. To take into account the data heterogeneity and its interpretation inherent uncertainty, we have chosen the Dempster/Shafer approach. As the classes of interest will form homogenous regions in the observed scene, the spatial context represents an additional source of information. To express that constraint of spatial homogeneity, we use the Markovian modelisation. This provides the joint probability distribution of the class labels. Next we express this distribution in terms of the Dempster/Shafer theory in order to fuse it with the observations given by the images.

A new decision criteria accelerates the convergence of the iterative classification procedure.

Une des possibilités est la théorie de l'évidence.

De part sa nature mathématique, son application aux problèmes réels pose encore des difficultés. Il s'agit surtout de l'estimation des masses élémentaires, un problème qui a souvent été traité de manière heuristique [1][2][3][4]. Nous utilisons dans la suite un jeu de masses qui a été déduit à partir d'un ensemble d'axiomes (décrit en 2.2) [5].

Pour obtenir des classifications contenant des zones homogènes, nous avons adopté le modèle markovien. Sa puissance a déjà été montré dans des nombreuses applications. Le problème est donc ici de transférer correctement les mesures probabilistes dans le cadre évidentiel.

2 FUSION D'IMAGES PAR LA THÉORIE DE DEMPSTER/SHAFFER

La théorie de Dempster/Shafer (ou, selon Shafer, la théorie de l'évidence) peut être considérée comme une généralisation de la théorie des probabilités. Elle est décrite en détail dans [6]. Une description brève, mais exhaustive, se trouve dans [5]. Dans la suite, nous ne présentons que les notions clé de cette théorie.

2.1 Les notions clé de la théorie de l'évidence

La masse élémentaire

Soit un ensemble donné $\Theta = \{\theta_1, \theta_2, \dots, \theta_N\}$. Les N événements (classes) θ_i sont exclusifs et exhaustifs. L'ensemble de tous ces événements représente le monde et s'appelle le cadre de discernement. Nous associons un degré de confiance à chaque



sous-ensemble $A \in 2^{\Theta}$, Ce degré de confiance sera appelé la masse élémentaire. Elle est définie de 2^{Θ} sur $[0,1]$ par:

$$m(\emptyset) = 0$$

$$\sum_{A \subset \Theta} m(A) = 1$$

Les éléments focaux sont les sous-ensembles dont la masse est non-nulle. Nous remarquons que l'approche de Dempster/Shafer se réduit sur la théorie des probabilités si les éléments focaux se réduisent sur les singletons θ_i .

La plausibilité

Sur les sous-ensembles, une fonction de plausibilité peut être définie par:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B)$$

Elle est la somme des masses de toutes les propositions qui contribuent à la proposition considérée. Elle peut être interprétée comme une mesure de vraisemblance maximale.

La combinaison des sources distinctes

Deux fonctions de masse m_i et m_j seront combinées selon la règle de Dempster:

$$(m_i \oplus m_j)(A) = K \sum_{X \cap Y = A} m_i(X) m_j(Y)$$

$$K^{-1} = 1 - \sum_{X \cap Y = \emptyset} m_i(X) m_j(Y)$$

K est une constante de normalisation. Elle peut être interprétée comme une mesure de conflit. Il est important de remarquer que l'application de cette règle requiert l'indépendance des sources fusionnées. La règle est commutative et associative, c'est-à-dire que l'ordre de combinaison des données n'a aucune influence sur le résultat.

2.2 Conversion des densités en masses

Appriou [5] a imposé trois propriétés initiales qui devront être remplies par les fonctions de masse:

1. Cohérence avec l'approche bayésienne dans le cas où les lois probabilistes sont parfaitement connus.
2. Expression des informations sous la forme des fonctions de support simple.
3. Cohérence avec l'association probabiliste des sources.

Un jeu de masses qui obéit à toutes les propriétés posées est, comme démontré par [5], le suivant:

$$m(\theta_i) = c_{ij} (1 - p(x|\theta_i) / R_j)$$

Le facteur c_{ij} indique le degré de correspondance avec la loi réelle. R_j est une constante de normalisation. Elle sera bornée par l'intervalle

$$R_j \in [0, \max_i p(x|\theta_i)^{-1}]$$

Maintenant, il est possible d'évaluer directement les plausibilités. Pour l'événement θ_i , elles s'estiment par:

$$Pl(\theta_i) = \prod_{j=1}^M (1 - c_{ij} + c_{ij} R_j p(x|\theta_i))$$

Il est facile de vérifier que pour $c_{ij}=1$ nous retrouvons le cas bayésien.

D'ailleurs, il est intéressant de considérer le cas des densités très faibles. Dans ce cas, l'observation considérée sera associée à la classe la moins sûre. Comme il est montré dans [5], le choix du facteur c_{ij} n'est pas très délicat. Pour des densités très vagues, il sera fixé à 0.9, pour des densités correspondant bien à la réalité, il sera fixé à 1.

3 INTÉGRATION DU CONTEXTE SPATIAL

Une classification aveugle ne donne généralement pas des résultats correspondants à la réalité, bien que le taux de bonne classification soit déjà bien élevé. Les classes à reconnaître se présenteront sous la forme de zones homogènes dans la vérité terrain. Pour obtenir le même effet dans l'image classifiée, il faut imposer des contraintes spatiales. Pour ceci, nous proposons l'utilisation d'une distribution de Gibbs [7][8]. Une application récente se trouve dans [9].

3.1 Le champs de Gibbs

Nous supposons que l'image de régions réelles Z représente un champ aléatoire markovien. Dû aux propriétés de ce type de champ, il se dégage les deux constatations suivantes:

1. Pour la prise en compte du contexte spatial, il suffit de se limiter au voisinage du pixel considéré.
2. Par rapport à ce voisinage, le champ aléatoire markovien doit suivre une distribution gibbsienne.

Une distribution gibbsienne est une mesure de probabilité de l'appartenance d'un pixel à une classe dérivé des propriétés de ses voisins. Elle a la forme:

$$P(\theta = \theta_k) = \frac{1}{Z} \exp \left\{ -\frac{1}{T} \sum_{C \in L} V_C(\theta_k) \right\}$$

Un sous-ensemble C est appelé une clique si C est soit un pixel isolé soit un groupe de pixels tel que chaque couple distinct de pixels est voisin d'un autre pour un voisinage donné. L'ensemble des cliques sera noté L . $V_C(\theta_k)$ est une fonction de potentiel associée à une clique C , T est une constante appelé température qui peut être fixée à 1, Z est une constante de normalisation définie par

$$Z = \sum_i \exp \left\{ -\frac{1}{T} \sum_{C \in L} V_C(\theta_i) \right\}$$

Dans la pratique, le champ aléatoire sera considéré comme homogène. Ceci implique que $V_C(\theta_k)$ ne dépend pas de la position de la clique dans le voisinage. Puisque les classes θ_i ne prennent que des valeurs discrètes, le champ aléatoire markovien peut être considéré comme un champ multi-logistique. Sa caractéristique principale est un paramètre positif associé à chaque type de clique. La clique d'un pixel isolé ne consiste qu'en ce pixel, et son potentiel est défini par

$$V_C(\theta) = \alpha_k \quad \text{if } \theta = \theta_k$$

Les cliques de deux pixels sont formées de pixel considéré et un pixel voisin. Leur potentiel est définie par

$$V_C(\theta) = \begin{cases} -\beta & \text{si tous les } \theta \in C \text{ sont identiques} \\ \beta & \text{sinon} \end{cases}$$

Les paramètres α_k et β sont positifs. Les paramètres α_k permettent de contrôler la fréquence d'apparition des classes correspondantes, le paramètre β détermine le degré d'homogénéité des régions. Par la suite, nous limitons le voisinage aux quatre plus proches voisins.

3.2 Conversion des probabilités a priori en masses

A l'aide d'une fonction de potentiel, nous avons pu exprimer les probabilités a priori en imposant le critère d'homogénéité spatiale. Pour la conversion des probabilités en masse, nous retenons la même expression qui nous sert pour la conversion des probabilités conditionnelles. En posant la constante de normalisation R_j à 1, nous obtenons:

$$m(\theta_i) = c_{ij}(1 - p(\theta_i))$$

Pour cela, l'évaluation des facteurs d'incertitude c_{ij} est nécessaire. Ils donnent un indice pour le degré de bonne correspondance avec la loi réelle. Pourtant, dans ce sens là une évaluation correcte de ces facteurs est difficile. Nous avons constaté que ces facteurs contrôlent les probabilités a priori globales des classes, et nous les posons dans ce sens. A priori, et pour les applications suivantes, nous les fixons égalent à 1.

3.3 La prise de décision à l'aide d'un nouveau critère

Lors de la fusion des masses m_f (décrivant les densités des classes de chaque image) et des masses m_p (décrivant le contexte spatial), le pixel considéré doit être classifié. Pour cela, nous avons choisi le critère du Maximum de plausibilité [5]. Pour optimiser la prise de décision, elle doit tenir compte de différentes contraintes (contexte spatial, énergie, etc.).

Pour cela, nous avons choisi un critère qui indique le degré de coïncidence d'une décision prise avec le vecteur m_f captant les degrés de confiance pour chaque classe a partir des observations. Un tel critère a été introduit par [10]. Il est de la forme

$$h(\theta_i) = \sum_{A \subset C, \theta_i \in A} \frac{m_f(A)}{|A|}$$

Maintenant, nous définissons un nouveau critère de décision qui est le produit de la plausibilité d'une décision et le degré de coïncidence avec le vecteur des mesures. Nous choisissons le classe θ_k tel que

$$k = \max_i \{h(\theta_i)Pl(\theta_i)\}$$

4 APPLICATION ET RÉSULTATS

4.1 Description des données

Nous avons créé une vérité terrain artificielle à l'aide d'un échantillonneur de Gibbs. Le modèle sous-jacent est le modèle MLL (multi logistic level) proposé par [7]. La vérité terrain est constituée de trois classes A (blanc), B (gris), C (gris foncé). Leur probabilités a priori sont fixées à $p(A)=0.44$, $p(B)=0.16$, $p(C)=0.39$. Le champ a été créé avec $\beta=2.0$ (voir figure 5).

Pour pouvoir bruite la vérité de terrain, les classes thématiques sont remplacés par une réponse spectrale idéale. Ensuite, deux images bruitées sont créées. A la première, un bruit gaussien additif a été superposé. Elle modélise une image issue d'un capteur optique. La deuxième a été bruitée avec un bruit gamma multiplicatif. Elle modélise une image issue d'un capteur radar

(deux vues). Au contraire de la première, elle mélange parfaitement les deux classes B et C.

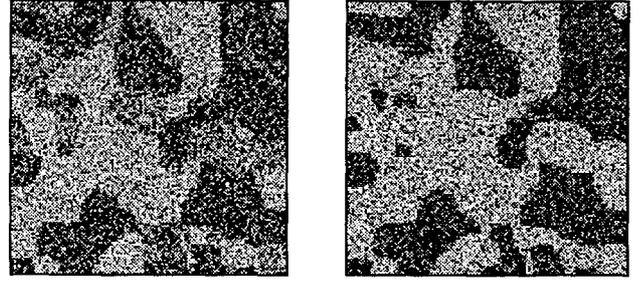


FIGURE 1. A gauche l'image bruitée gaussienne. Les densités des classes ont pour moyennes $\mu_A=80$, $\mu_B=100$, $\mu_C=120$ et un écart-type commun $\sigma=25$. A droite l'image bruitée gamma (deux vues). Les densités des classes ont les moyennes $\mu_A=20$, $\mu_{B \cup C}=80$.

4.2 Fusion et classification aveugle

Par la suite, nous supposons parfaitement connue les densités des classes. Nous avons posé les facteurs d'affaiblissement c_{ij} respectifs égal à 1. Le critère de décision est le critère du Maximum de plausibilité. Ceci revient à une classification par Maximum de Vraisemblance. Les résultats sont répertoriés dans le tableau 1.

Il est possible de constater que l'intégration des probabilités a priori globales augmente légèrement le taux de bonne classification globale. Pourtant, la probabilité d'apparition de la classe B s'atténue considérablement. Dans le cas présent, il n'est pas conseillé d'en tenir compte. L'application de l'algorithme pour l'amélioration du contexte spatial supprimerait celle-ci entièrement au bout de peu d'itérations.

source	A	B	C	global
f _{gauss}	67	31	66	61
+f _{gamma}	87	49	62	71
+p _{glob.}	91	14	83	75

TABLEAU 1. Taux de bonne classification pour une fusion des différentes sources.

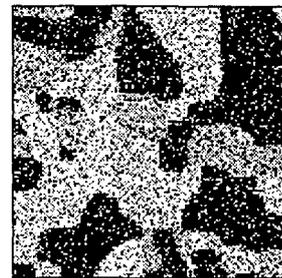


FIGURE 2. Classification aveugle lorsqu'une fusion des densités des images gaussienne et gamma.

4.3 Intégration du contexte spatial

Pour la conversion des probabilités a priori locales, les facteurs d'affaiblissement ont été fixés à 1. Un choix de valeurs plus faibles dégrade le résultat pour le cas présent. Ensuite, il faut poser les paramètres du champ de Markov. Ici, nous avons posé les α_k égaux à 0. Quant aux β , nous avons étudié plusieurs valeurs (tableau 2). Pour les valeurs égales ou supérieures à 1, l'algorithme



converge vers la classification "correcte". Le résultat de la classification au bout de 10 itérations est donnée dans la figure 4 et par le tableau 3. Il faut surtout retenir la bonne détection de la classe B. Puisque l'image gamma ne distingue pas les classes B et C, il y reste un certain conflit pour leur reconnaissance (matrice de confusion dans la figure 5).

Nous pouvons observer une convergence plus rapide avec l'utilisation d'un nouveau critère de décision (figure 3).

β	0.5	1.0	2.0	3.0	4.0
$t_{global} (\%)$	72	89	94	93	95

TABLEAU 2. Taux global de bonne classification en fonction de la valeur β (critère de décision: Maximum de plausibilité, 10 itérations).

	A	B	C	global
t (%)	99	80	98	95

TABLEAU 3. Taux de bonne classification au bout de 10 itérations (critère de décision: Maximum du produit plausibilité*h).

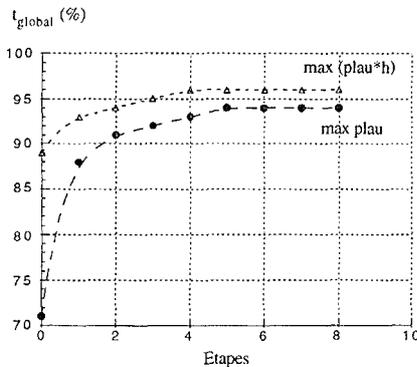


FIGURE 3. Taux de bonne classification globale pour deux critères de décision différents.

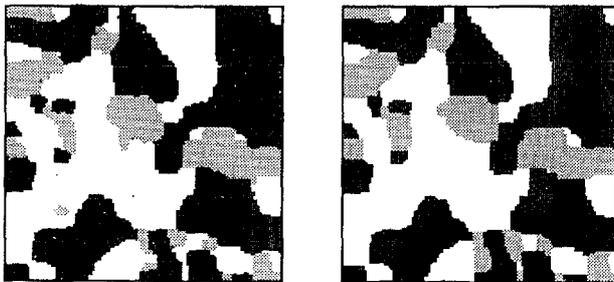
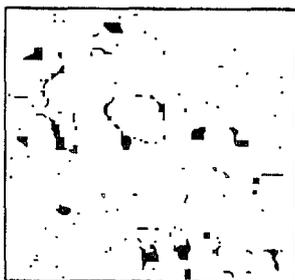


FIGURE 4. A gauche: la vérité terrain. A droite: classification intégrant le contexte spatial au bout de 10 étapes.



	A	B	C
a	7190	53	38
b	24	2215	120
c	27	431	6286

FIGURE 5. La carte d'erreur par rapport à la vérité terrain et la matrice de confusion pour la classification au bout de 10 étapes (A, B, C: classes de la vérité terrain; a, b, c: classification).

5 CONCLUSIONS ET PERSPECTIVES

Nous avons montré une possibilité de classifier un jeu d'images hétérogènes sous la contrainte de l'homogénéité spatiale dans le cadre de la théorie de Dempster/Shafér. La modélisation markovienne de la contrainte spatiale améliore, aussi dans le cadre évidentiel, nettement la classification. Néanmoins, la performance de la procédure itérative dépend, évidemment, de la fiabilité de l'initialisation du champ de régions, lequel est donné ici par une classification Maximum de Vraisemblance préliminaire.

Le nouveau critère de décision accélère nettement la convergence de l'algorithme.

Les taux de bonne classification très élevés dans le cas d'école considéré sont dû à l'utilisation de paramètres théoriques et à la simplicité du terrain.

Une étude plus exhaustive doit être menée sur les paramètres décrivant les probabilités a priori déduit par la modélisation markovienne du champ de régions.

Le modèle ne devrait pas se limiter à l'association des potentiels aux classes θ_i , mais devrait permettre la prise en compte de tous les sous-ensembles du cadre de discernement.

La variation de la température permet une classification qui suit l'idée de l'algorithme de recuit simulé.

Enfin, une estimation plus élaborée du facteur d'affaiblissement c_{ij} pour exprimer les probabilités a priori sous la forme d'une fonction de masse est possible. Elle devrait encore améliorer la classification.

6 RÉFÉRENCES

- [1] T. Lee, J.A. Richards, and P.H. Swain. Probabilistic and evidential approaches for multisource data analysis. In *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-25, No. 3, May 1987, pp. 283-292.
- [2] H. Rasoulin, L.F. Kazda, W.E. Thompson, R. Parra-Loera. Application of the mathematical theory of evidence to the image cueing and image segmentation problem. *SPIE Vol. 1310 Signal and Image Processing Systems Performance Evaluation*. 1990, pp. 199-206.
- [3] R.J. Safranek, S. Gottschlich, A.C. Kak. Evidence Accumulation using binary frames of discernment for verification vision. *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 4, August 1990.
- [4] E. Mandler, J. Schürmann. *Combining the classification results of independant classifiers based on the Dempster/Shafér theory of evidence*. In *Pattern Recognition and Artificial Intelligence*, E.S. Gelsema and L.N. Kanal (Editors), Elsevier Science Publishers B.V. (North Holland), 1988.
- [5] A. Appriou. Probabilités et incertitude en fusion de données multi-senseurs. *Revue Scientifique et Technique de la défense*. 1er trimestre 1991, pp. 27-40.
- [6] G. Shafer. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press, 1976.
- [7] S. Lakshman, H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-11, No. 8, August 1989, pp. 799-813.
- [8] S. Geman, D. Geman. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, No. 6, 1984, pp. 721-741.
- [9] D. Zhang, L. Van Gool, A. Oosterlinck. Classification of remote sensing images with the aid of Gibbs distribution. *SPIE Vol. 2315*, pp. 462-471, 1994.
- [10] D.Y. Suh, R.L. Eisner, R.M. Merserau, R.I. Pettigrew. Knowledge-based system for boundary detection of four dimensional cardiac magnetic resonance image sequences, *IEEE Transactions on Medical Imaging*, Vol. 12, No. 1, pp. 65-72, March 1993.