



Évaluation des performances du modèle bayésien de fusion appliqué à l'imagerie satellitaire

Stéphane Chauvin

Département Images, École Nationale des Télécommunications,

46 rue Barrault, 75634 Paris Cedex 13, France

Tél : (1) 45 81 72 18 - Fax : (1) 45 81 37 94 - Email : chauvin@ima.enst.fr

RÉSUMÉ

ABSTRACT

Nous présentons dans cet article différents modèles de fusion d'images à bas niveau par l'inférence bayésienne en vue de la classification de scènes satellitaires. Le processus se décompose en une étape d'apprentissage pour laquelle se posent les problèmes du choix des densités de probabilité conditionnelles et d'estimation des probabilités a priori, une étape de combinaison des probabilités pour laquelle on évalue l'hypothèse de dépendance entre les images contre l'hypothèse d'indépendance et une étape de décision pour laquelle on définit une mesure de qualité des images et une classe de rejet dont on estime de manière non supervisée la probabilité.

We present in this paper a Bayesian inference method for numerical data fusion dedicated to satellite image classification. The fusion process consists in learning phase, a combination phase and a decision phase. During the learning phase, the conditional probability densities are evaluated, and the prior class probabilities is estimated. During the combination phase, the hypothesis of dependence between the images is tested against the hypothesis of independence. During the decision phase, we estimate a parameter indicating image quality as well as a the probability for the dismiss class.

1 Introduction

L'objectif de la fusion de données est de faire a synthèse de plusieurs informations pour obtenir par exemple de meilleures décisions et favoriser la coopération de p capteurs présents dans un système. Les modèles probabilistes de fusion de données suggérés dans la littérature imposent des hypothèses fortes sur les étapes du processus (densités gaussiennes, hypothèses d'indépendance entre les capteurs, [LUO-89], [LEE-87]) et nécessitent généralement une connaissance particulière et spécifique du système pour choisir, en fonction de la classe recherchée, le ou les capteurs nécessaires parmi ceux disponibles (par exemple HOUZ-91] et [MANG-93]). Notre étude propose, pour chacune des étapes, une série de tests statistiques permettant de vérifier les hypothèses sur les éléments intervenant dans la formule de Bayes et fournit une méthodologie pour la construction du processus de fusion bayésien.

Dans la suite, on note n le nombre de classes à détecter et $C = \{C_1, C_2, \dots, C_n\}$ l'ensemble contenant toutes les classes C_i , $i \in \{1, \dots, n\}$ (espace

de décision). Ω est l'espace des niveaux de gris des p images à combiner. $\omega = \{\omega_1, \dots, \omega_p\} \in \Omega$ est le vecteur des niveaux de gris d'un point $s \in S$ de la scène. On rappelle l'équation de Bayes $\forall i \in \{1, \dots, n\} \forall \omega \in \Omega$:

$$P(C_i/\omega_1, \dots, \omega_p) = \frac{P(\omega_1, \dots, \omega_p/C_i)}{P(\omega_1, \dots, \omega_p)} P(C_i).$$

Le cas pratique présenté dans cet article est dédié à la détection des villes, des fleuves et de la classe $C_3 = \overline{C_{ville}} \cap \overline{C_{fleuve}}$ à partir d'images du satellite SPOT multispectral (XS1, XS2, XS3) sur la région de Vignola en Italie.

2 Apprentissage

La formule de Bayes met en jeu deux probabilités qui font l'objet d'une estimation : probabilités conditionnelles $P(\omega_j/C_i)$ (fréquence d'apparition du niveau de gris ω_j dans la classe C_i sur l'image j) et les probabilités a priori $P(C_i)$ de chaque classe (occurrence des classes sur la scène).

2.1 Probabilités conditionnelles

Les mesures statistiques du kurtosis γ_2 (rapport



du moment statistique d'ordre quatre sur le carré de la variance) et du coefficient d'asymétrie γ_1 (rapport du moment statistique d'ordre trois sur l'écart type au cube) permettent d'émettre des hypothèses sur les lois sous-jacentes et de dégager des comportements significatifs [SAPO-78].

Sur le canal 1 de SPOT, les mesures du kurtosis et du coefficient d'asymétrie de la classe ville calculées sur deux zones d'apprentissage montrent d'une part que les lois sous-jacentes sont asymétriques ($\gamma_1 \neq 0$) et indiquent d'autre part qu'il y a un fort regroupement sur un ensemble réduit de niveaux de gris ($\gamma_2 \simeq 3$) proche de l'hypothèse gaussienne (tableau 1). Dans ce cas, les tests

Zone 1		Zone 2	
γ_2	γ_1	γ_2	γ_1
3,57	0,25	2,8	0,29

TAB. 1 - : *Mesure du kurtosis et du coefficient d'asymétrie de la classe ville dans XS1.*

statistiques d'égalité des moyennes de Student et d'égalité des variances de Fisher dénoncent l'instabilité des estimations des paramètres selon les zones d'apprentissage et par conséquent réfutent l'hypothèse d'une loi gaussienne. Les conclusions analogues sont obtenues pour d'autres lois paramétriques et nous conduisent à estimer ces probabilités de manière non-paramétrique, directement par l'histogramme normalisé des niveaux de gris.

La confiance que l'on a sur l'estimation \hat{P}_i obtenue sur l'histogramme de la probabilité conditionnelle P_i d'appartenance d'un point de niveau ω à la classe C_i est quantifiée à partir de la loi normale de moyenne P_i et de variance $P_i(1 - P_i)/nb_i$ dans le cas où la population totale de points sur l'image appartenant à la classe C_i est approximativement cinq fois supérieure à la taille de l'échantillon nb_i . Dans le cas contraire, le calcul d'erreur se fait à partir de la loi binômiale [SAPO-78]. L'intervalle de confiance qui résulte de la loi gaussienne donne une faible confiance sur l'ensemble des probabilités conditionnelles \hat{P}_i estimées sur les images de Vignola, lorsque le nombre de points des échantillons d'une classe est inférieur à 10000. Les erreurs d'estimation affectent essentiellement les niveaux de gris peu présents sur les images (généralement les contours des classes). L'apprentissage des probabilités conditionnelles sur de petites images et a fortiori pour des classes peu présentes sur la scène satellitaire, est un cas limite d'utilisation de l'inférence bayésienne.

2.2 Probabilités a priori

La principale critique faite dans la littérature à l'inférence bayésienne pour la classification porte sur la difficulté d'estimer les probabilités a priori. Sans cette information, il est d'une part impossible de rendre le processus de fusion automatique (il faut qu'un expert livre au système les occurrences de chaque classe sur chaque scène), et d'autre part si les probabilités a priori sont prises toutes égales, les erreurs de décision par la règle du maximum a posteriori sont importantes [DUDA-73]. Nous proposons une estimation des probabilités a priori sous des hypothèses que nous présentons tout d'abord.

Nous supposons ici qu'il existe des réalisations $\omega \in \Omega' \subset \Omega$ pour lesquelles il est sûr que ces points appartiennent à une et une seule classe. Pour de telles réalisations, on déduit la probabilité a priori :

$$\forall i \in \{1, \dots, n\}, \exists \omega \in \Omega', P(\omega/C_k) = 0, \forall k \neq i$$

$$\Rightarrow P(C_i/\omega) = 1 \Rightarrow P(C_i) = \frac{P(\omega)}{P(\omega/C_i)}. \quad (1)$$

Cette solution suppose qu'il existe pour chaque classe une image ayant des points que l'on peut classer sans ambiguïté ni incertitude. Nous visualisons sur les graphiques de la figure 1 les valeurs conjointes des probabilités $P(\omega)$ et des probabilités conditionnelles $P(\omega/ville)$. On distingue deux populations, celle pour laquelle le rapport $P(\omega)/P(\omega/ville)$ est sensiblement constant et vérifie l'équation 1, et celle pour laquelle ce rapport n'est pas constant. La constante ainsi visualisée correspond à la probabilité a priori $P(ville) \simeq 19\%$. L'algorithme proposé, résultant de l'équation 1 repose sur la moyenne des valeurs sur tout Ω' . $P(fleuve)$ a été estimée à 1,87% et $P(C_3)$ à 77,18% sur le canal 3, ce qui correspond à la réalité de la scène.

3 Fusion

L'opérateur de fusion bayésien a la propriété d'être associatif et conjonctif. La modélisation du processus de fusion bayésien peut être statique (informations disponibles simultanément et on estime alors directement $P(\omega_i, \dots, \omega_p/C_i)$) ou dynamique (informations disponibles successivement et on utilise le conditionnement bayésien faisant apparaître les probabilités conditionnelles aux informations $\prod_{j=1}^p P(\omega_j / \bigcap_{k=1}^{j-1} \omega_k, C_i)$). Sous l'hypothèse d'indépendance entre les images, l'étape de fusion est réduite à $P(\omega/C_i) = \prod_{j=1}^p P(\omega_j/C_i)$

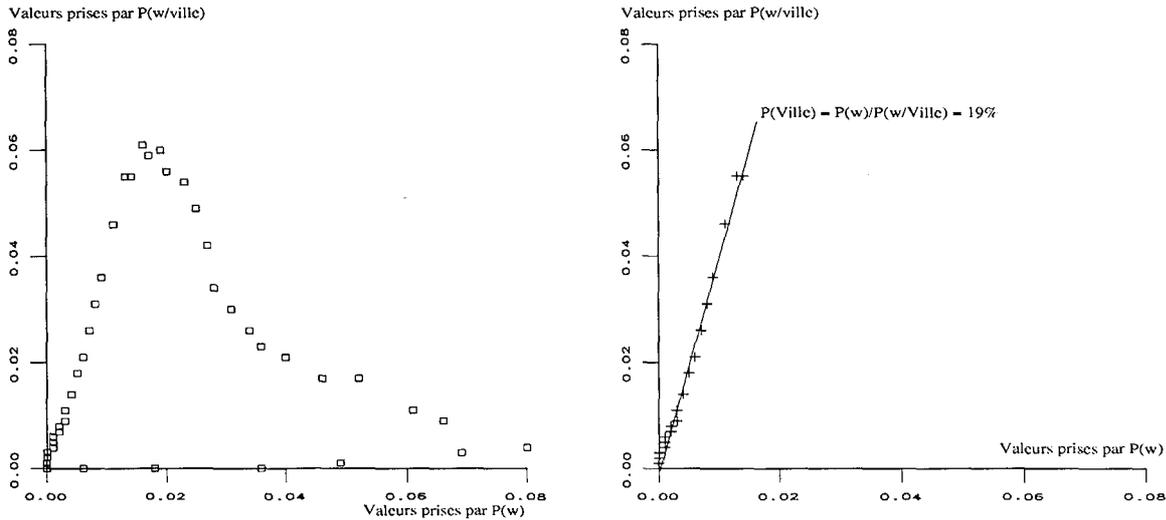


FIG. 1 - : Valeurs conjointes des probabilités conditionnelles $P(\omega/ville)$ en abscisse et $P(\omega)$ en ordonnée sur le canal XS1, à gauche population totale Ω et à droite population Ω . La pente de la droite représente la probabilité a priori $P(ville) \simeq 19\%$. Le passage de la courbe de gauche à la courbe de droite s'effectue par la sélection des points tels que $P(\omega/fleuve) = P(\omega/C_3) = 0$.

(produit des probabilités marginales). On se propose de tester l'hypothèse de dépendance représentée par la probabilité jointe $P(\omega/C_i)$ contre l'hypothèse d'indépendance représentée par le produit $\prod_{i=1}^p P(\omega_j/C_i)$ en calculant la distance du χ^2 (écart à l'indépendance) et qui s'écrit $\forall i \in \{1, \dots, n\}$:

$$d^2 = \sum_{\omega \in \Omega} \frac{(P(\omega/C_i) - \prod_{i=1}^p P(\omega_j))^2}{\prod_{i=1}^p P(\omega_j)}$$

d^2 suit une loi du χ^2_ν avec ν le degré de liberté égal au nombre de points de Ω pour lesquelles les probabilités ne sont pas nulles. L'hypothèse d'indépendance entre les images est acceptée si d^2 est inférieure à la valeur théorique χ^2 au seuil de probabilité α (écart maximum autorisé). La borne $d^2 = \sqrt{\nu}$ sera atteinte lors de la dépendance fonctionnelle [SAPO-78]. Cette mesure est équivalente à la fonction de lien [MARC-79] et de redondance (ou entropie) de Kullback-Leiber sous l'hypothèse d'indépendance [DEV-87]. On constate sur les images SPOT que l'hypothèse d'indépendance entre les images est justifiée (la mesure d^2 calculée pour deux images est voisine de 0) mais qu'elle est moins significative entre les images conditionnellement aux classes (la mesure entre les deux premiers canaux SPOT conditionnellement à la classe ville est égale à $12,72 \leq 82,80 = \chi^2$). Les tests effectués ont montré qu'au-delà de deux images, l'hypothèse d'indépendance l'emporte sur l'hypothèse de dépendance des images quel que soit le conditionnement.

La mesure d^2 permet ainsi systématiquement de vérifier et d'évaluer l'hypothèse de dépendance entre les images contre l'hypothèse forte d'indépendance.

4 Décision

La probabilité d'erreur de décision d'un système de classification à n classes par la règle du maximum des probabilités a posteriori s'écrit :

$$P(\text{Erreur}) = 1 - \sum_{\omega \in \Omega} \max_{i=1}^n P(\omega/C_i) \cdot P(C_i). \quad (2)$$

Elle est nulle lorsque les supports des probabilités conditionnelles sont disjoints. On estime alors la qualité d'une image pour une classe en fonction de la séparabilité des supports des probabilités.

Qualité des images : on utilise un critère statistique fondé sur l'inertie intra-classes de Fisher notée $F(C_i)$ pour une classe C_i [GOOD-79], qui s'écrit :

$$F(C_i) = \frac{(\mu_i - \mu_{c(i)})^2}{P(C_i)\sigma_i^2 + (1 - P(C_i)) \cdot \sigma_{c(i)}^2}$$

avec μ_i la moyenne et σ_i l'écart type des niveaux de gris de la classe C_i ($\mu_{c(i)}$ et $\sigma_{c(i)}$ sont les paramètres de la classe complémentaire \bar{C}_i calculés sur les zones d'apprentissage des autres classes). L'inertie de la classe C_i est élevée si elle est discriminée des autres classes. Le tableau 2 du calcul des inerties sur les images SPOT des classes ville, fleuve et C_3 montre que le canal 3 discrimine le mieux les informations des trois classes. Lorsque



le nombre d'images est important, cette mesure permet de sélectionner les images de meilleure qualité.

$F(C_i)$	Ville (%)	Fleuve (%)	C_3 (%)
XS1	1,87	9,79	1,95
XS2	6,08	10,09	6,09
XS3	6,36	22,61	8,26

TAB. 2 - : Inertie des trois canaux SPOT conditionnellement aux trois classes.

Calcul d'une probabilité de rejet : la somme des probabilités a priori estimées précédemment est inférieure à 1 s'il existe une proportion de niveaux de gris appartenant à une classe $C_{n+1} = \bigcap_{i=1}^n \bar{C}_i$ non définie. La décision du maximum des probabilités a posteriori porte alors sur $n+1$ classes. L'estimation de la probabilité a posteriori de la classe C_{n+1} (considérée comme une classe de rejet) est calculée à partir des mesures sur les autres classes et s'écrit :

$$P(C_{n+1}/\omega) = 1 - \sum_{i=1}^n \frac{P(\omega/C_i)}{P(\omega)} P(C_i).$$

5 Résultats et conclusion

Le tableau 3 présente les résultats des probabilités d'erreur de décision (équation 2) après agrégation des trois canaux SPOT : probabilités a priori toutes égales, probabilités a priori exactes et lorsque la classe de rejet est mise en jeu avec les probabilités a priori exactes (la probabilité d'erreur de décision est minimum). L'image 2 correspond à la détection des villes, des fleuves et $C_3 \cup \{\text{classe de rejet}\}$. Les performances du modèle bayésien

$P(C_i) = 1/n$	$P(C_i)$	$P(C_i) + \text{Rejet}$
11,22%	7,91%	6,23%

TAB. 3 - : Probabilité d'erreur de décision des trois modèles de fusion.

sont fonctions d'une part de l'apprentissage (la taille des zones doit être grande et supérieure à 10000 points pour chaque classe), du choix de la densité de probabilité conditionnelle et de l'estimation des probabilités a priori, d'autre part de l'hypothèse de dépendance entre les images (le test du χ^2 mesure la validité des hypothèses) et en dernier lieu la qualité des images. Le facteur de qualité permet de combiner les images en commençant par les meilleures et de s'arrêter lorsque la probabilité d'erreur est suffisamment petite.

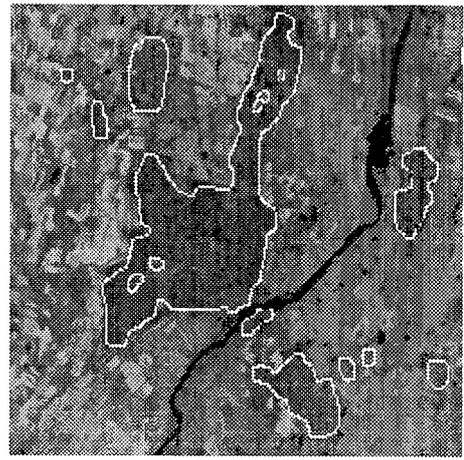


FIG. 2 - : Détail de l'image de la détection des villes, des fleuves et $C_3 \cup \{\text{classe de rejet}\}$ issue de la fusion des trois canaux du satellite SPOT.

L'objectif de nos travaux actuels est d'améliorer les performances des modèles statistiques en utilisant dans le cadre de la théorie des ensembles flous, des opérateurs de combinaison différents du produit quand l'hypothèse d'indépendance n'est pas vérifiée.

Références

- [DEV-87] L. DEVROYE, *A Course In Density Estimation*, Birkhäuser, Stuttgart, 1987.
- [DUDA-73] R. DUDA ET P. HART, *Pattern Classification and Scene Analysis*, Wiley Interscience, John Wiley et Sons, Inc, New York, 1973.
- [GOOD-79] L. A. GOODMAN ET W. H. KRUSKAL, *Measures of Association for Cross Classifications*, Springer Series in Statistics, Berlin, Springer Verlag ed., 1979.
- [HOUZ-91] S. HOUZELLE ET G. GIRAUDON, *Extraction automatique d'objets par fusion de données d'images satellites*, AFCET, Lyon, (1991), pp. 467-475.
- [LEE-87] T. LEE, J. A. RICHARDS, ET P. H. SWAIN, *Probabilistic and evidential approaches for multisources data analysis*, IEEE Transactions on Geoscience and Remote Sensing, GE-25 (1987), pp. 283-293.
- [LUO-89] R. C. LUO ET M. G. KAY, *Multisensor integration and fusion in intelligent systems*, IEEE Transactions on Systems, Man, and Cybernetics, 19 (1989), pp. 901-931.
- [MANG-93] M. MANGOLINI, T. RANCHIN, ET L. WALD, *Fusion d'images SPOT Multispectrale (XS) et Panchromatique (P), et d'images Radar*, Congrès SPOT/ERS, (1993), pp. 1-9.
- [MARC-79] J. F. MARCOTORCINO ET P. MICHAUD, *Optimisation en analyse ordinaire des données*, MASSON, Paris, 1979.
- [SAPO-78] G. SAPORTA, *Théories et méthodes statistiques*, Princeton University Press, 1978.