

# Quelques améliorations à la segmentation d'images bayésienne

## Seconde partie : classification

---

## Some Improvements to Bayesian Image Segmentation Part Two : Classification

par Xavier DESCOMBES, Robin MORRIS, Josiane ZÉRUBIA

INRIA, 2004, route des Lucioles, BP 93  
06902 Sophia Antipolis, Cedex  
email : Prénom.Nom@inria.fr

### *résumé et mots clés*

Nous nous plaçons dans le cadre de la classification automatique. Nous abordons le problème de l'estimation du nombre de classes et des paramètres qui leurs sont associés. Nous proposons une méthode utilisant l'hypothèse contextuelle inhérente aux images pour discriminer les différentes classes. Cette méthode est validée à la fois sur le plan théorique et sur des images de synthèse et des images réelles. Nous montrons, en outre, que la méthode proposée a un domaine de validité plus étendu que les méthodes fondées sur une analyse des modes de l'histogramme. Nous discutons ensuite de la forme du potentiel d'attache aux données dérivé de cette classification dans le cadre d'une segmentation markovienne. Les résultats sont obtenus avec deux modèles *a priori* différents : le modèle de Potts et le chien-modèle.

Segmentation d'images, Champs de Markov, estimation, mélange de distributions.

### *abstract and key words*

We consider the automatic classification framework, using a Markov model. We address the problem of estimating the number of classes and the associated class parameters. We propose a method using the contextual information inherent in images to discriminate different classes in the case of mixture distributions with strongly mixed classes. This method is validated theoretically and practically, using synthetical images and real data. We prove that the proposed method has a validity domain larger than the methods based on a histogram analysis. We then discuss the shape of the data driven potential induced by the detected classes in a Markovian framework. Results are obtained by using two priors : the Potts model and the Chien-model.

Image segmentation, Markov Random Fields, estimation, mixtures.

## 1. Introduction

Une étape primordiale pour obtenir des algorithmes de segmentation non supervisés concerne l'estimation des paramètres associés aux différentes classes. L'utilisation du contexte bayésien et de modèles *a priori* suppose la connaissance des différentes classes et au moins une initialisation des paramètres qui les définissent, [11], [9]. Nous abordons dans cet article le problème de la classification automatique pour laquelle un label est associé à chaque pixel sans contenu sémantique. Nous supposons que

cette classification est uniquement déterminée par le niveau de gris des données, c'est-à-dire que nous ne considérons pas de paramètres texturaux. Nous nous plaçons dans le cadre de la classification non supervisée ou « clustering ». Une revue des différentes approches est présentée dans [13]. Bien que prometteurs pour certaines applications, les algorithmes de classification automatique comme l'algorithme des K-moyennes ou l'algorithme des C-moyennes floues nécessitent une bonne initialisation du nombre et du centre des classes. Dans [12], une analyse dans un espace d'échelle sur l'histogramme de l'image est utilisée pour déterminer le nombre des classes ainsi que leurs centres dans le cas d'un mélange de distributions gaussiennes. Cette méthode est

donc essentiellement fondée sur l'analyse de la multi-modalité de l'histogramme. Aussi, lorsque les centres de deux classes consécutives sont proches par rapport à la variance des classes, les deux classes ne sont pas discriminées [14]. Nous proposons donc d'utiliser une propriété contextuelle inhérente aux images permettant de discriminer les différentes composantes de distributions fortement mélangées. Nous calculons l'histogramme de la variance des pixels conditionnellement à la valeur de leurs voisins et utilisons ensuite une analyse dans un espace d'échelle. Outre un algorithme de classification automatique, cette procédure nous permet de définir le terme d'attache aux données d'un modèle markovien.

Dans le paragraphe 2, nous explicitons la méthode proposée. Nous montrons ensuite théoriquement l'apport de l'hypothèse contextuelle en comparant les domaines de validité de la méthode proposée et de toute méthode fondée sur une analyse de la multi-modalité de l'histogramme. Le paragraphe 4 est consacré à l'algorithme de segmentation utilisant l'estimation des classes et un modèle *a priori* markovien. Des résultats sont présentés sur une image de synthèse et sur des données réelles. Le paragraphe 5 compare différentes formes de l'attache aux données dans le cadre d'une modélisation markovienne. Pour finir, nous concluons et introduisons nos dernières recherches en vue d'obtenir un algorithme totalement non supervisé.

## 2. méthode de l'Histogramme de la Variance Conditionnelle (HVC)

Nous proposons dans ce paragraphe une méthode de classification automatique. À partir des données, nous extrayons le nombre de classes et les paramètres qui leur sont associés. Ces paramètres sont par la suite utilisés pour définir le terme d'attache aux données d'une modélisation markovienne. Plusieurs méthodes d'estimation des paramètres d'une loi *a priori* markovienne ont été proposées. Les plus répandues sont les méthodes de codages [1] pour des topologies particulières ou bien celles utilisant un critère de pseudo-maximum de vraisemblance dans un cadre plus général [2], [9], [7]. Nous étudions actuellement une méthode fondée sur un critère de maximum de vraisemblance ("Markov Chain Monte Carlo Maximum Likelihood", [21]). Le problème de la classification est différent et plus général car non uniquement lié à la modélisation markovienne.

Nous supposons que les différentes classes sont définies à partir des niveaux de gris uniquement, c'est-à-dire que nous ne tenons pas compte de critères texturaux. Nous considérons également que nous avons une seule image de données. Le problème de l'estimation du nombre des classes se ramène à l'estimation des composantes d'un mélange de distributions.

Une première approche consiste à rechercher les modes de l'histogramme de l'image. Néanmoins, cette approche peut conduire

à une mauvaise classification dans le cas de distributions fortement mélangées. Supposons en effet que deux classes de l'image aient des moyennes proches l'une de l'autre, comparativement à leurs variances. Ces distributions seront alors représentées par un unique mode dans l'histogramme. Dans ce cas, un algorithme de détection de modes ne séparera pas ces deux classes. Nous montrons que l'utilisation d'une variance conditionnelle (variance de la loi des voisins d'un pixel conditionnellement à sa valeur) permet une meilleure séparation des modes.

### 2.1. principe

Considérons un mélange de deux distributions  $P_1$  and  $P_2$ , dont les moyennes (resp. variances) sont notées  $\mu_1$  et  $\mu_2$  (resp.  $\sigma_1$  et  $\sigma_2$ ). Supposons que la probabilité de la classe 1 soit donnée par  $\pi$  et, par conséquent, la probabilité de la classe 2 par  $1 - \pi$ . Si ces distributions sont suffisamment proches, l'histogramme résultant sera alors unimodal (voir figure 1.a). En utilisant une hypothèse contextuelle, nous pouvons retrouver les deux distributions correspondant aux deux classes ( $C_1$  et  $C_2$ ). Notons que nous ne considérons pas connaître la loi des différentes distributions. En effet, si nous ajoutons l'hypothèse gaussienne par exemple, les deux distributions peuvent être retrouvées par un algorithme de mise en correspondance d'une somme de gaussiennes avec un histogramme. Dans la pratique, nous ne connaissons que très rarement la forme des distributions des classes. Nous pallions donc ce manque de connaissance par une hypothèse contextuelle, valide lorsque l'on considère des images comme données. Nous effectuons donc l'hypothèse suivante :

**Hypothèse 1** *Les voisins d'un pixel appartiennent à la même classe que ce pixel.*

Cette hypothèse est trivialement vérifiée pour tous les pixels n'appartenant pas aux bords d'une région.

Considérons un pixel  $s$  dont le niveau de gris  $\omega_s$  est égal à  $x$ . Par application de la loi de Bayes, la probabilité que ce pixel appartienne à la classe 1 (resp. 2) est donnée par :

$$P(s \in C_1 | \omega_s = x) = \frac{P(\omega_s = x | s \in C_1)P(s \in C_1)}{P(\omega_s = x | s \in C_1)P(s \in C_1) + P(\omega_s = x | s \in C_2)P(s \in C_2)},$$

$$P(s \in C_2 | \omega_s = x) = \frac{P(\omega_s = x | s \in C_2)P(s \in C_2)}{P(\omega_s = x | s \in C_1)P(s \in C_1) + P(\omega_s = x | s \in C_2)P(s \in C_2)}, \quad (1)$$

ce qui peut encore s'écrire :

$$A_1(x) = P(s \in C_1 | \omega_s = x) = \frac{\pi P_1(x)}{\pi P_1(x) + (1 - \pi)P_2(x)},$$

$$A_2(x) = P(s \in C_2 | \omega_s = x) = \frac{(1 - \pi)P_2(x)}{\pi P_1(x) + (1 - \pi)P_2(x)}. \quad (2)$$

Utilisant l'hypothèse 1,  $A_1(x)$  (resp.  $A_2(x)$ ) représente la probabilité qu'un voisin de  $s$  appartienne à la classe 1 (resp. classe 2). Considérons la variable aléatoire  $\Lambda(x)$ , espérance conditionnelle (au sens de Kolmogorov) de la moyenne empirique des niveaux

de gris des sites voisins  $r$  d'un site donné  $s$  sachant que le niveau de gris de celui-ci vaut  $x$  :

$$\Lambda(x) = E \left( \frac{1}{|N_s|} \sum_{r \in N_s} \omega_r \middle| \omega_s = x \right). \quad (3)$$

En supposant les sites  $r$  indépendants conditionnellement à  $s$ , nous avons :

$$\begin{aligned} \Lambda(x) &= \frac{1}{|N_s|} \sum_{r \in N_s} E(\omega_r | \omega_s = x) \\ &= \frac{1}{|N_s|} \sum_{r \in N_s} \int \xi P(\omega_r = \xi | \omega_s = x) d\xi. \end{aligned} \quad (4)$$

(5)

D'autre part comme nous avons considéré deux classes :

$$\begin{aligned} P(\omega_r = \xi | \omega_s = x) &= P(\omega_r = \xi, s \in C_1 | \omega_s = x) \\ &\quad + P(\omega_r = \xi, s \in C_2 | \omega_s = x), \end{aligned} \quad (6)$$

d'où en appliquant la règle de Bayes :

$$\begin{aligned} P(\omega_r = \xi | \omega_s = x) &= P(\omega_r = \xi | \omega_s = x, s \in C_1) P(s \in C_1 | \omega_s = x) \\ &\quad + P(\omega_r = \xi | \omega_s = x, s \in C_2) P(s \in C_2 | \omega_s = x), \\ &= P(\omega_r = \xi | \omega_s = x, s \in C_1) A_1(x) \\ &\quad + P(\omega_r = \xi | \omega_s = x, s \in C_2) A_2(x). \end{aligned} \quad (7)$$

En utilisant l'indépendance conditionnelle des sites  $r$  et  $s$  ainsi que l'hypothèse 1, nous obtenons :

$$P(\omega_r = \xi | \omega_s = x) = P_1(\xi) A_1(x) + P_2(\xi) A_2(x), \quad (8)$$

où  $P_1$  et  $P_2$  sont les lois des classes  $C_1$  et  $C_2$ .

Ceci nous permet d'écrire pour  $\Lambda(x)$  :

$$\begin{aligned} \Lambda(x) &= \frac{1}{|N_s|} \sum_{r \in N_s} \int \xi (P_1(\xi) A_1(x) + P_2(\xi) A_2(x)) d\xi, \\ &= \left( \int \xi P_1(\xi) d\xi \right) A_1(x) + \left( \int \xi P_2(\xi) d\xi \right) A_2(x). \end{aligned} \quad (9)$$

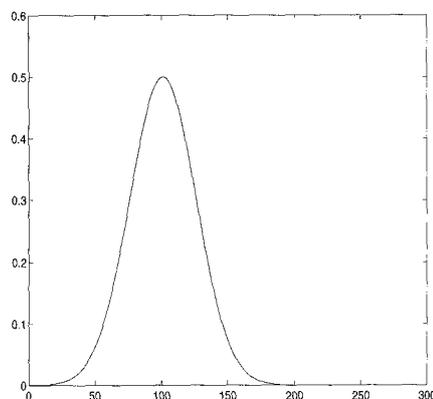
et par conséquent :

$$\Lambda(x) = A_1(x) \mu_1 + A_2(x) \mu_2. \quad (10)$$

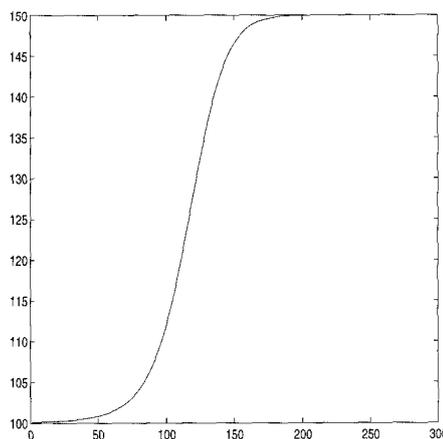
Si tous les pixels de niveau de gris égal à  $x$  appartiennent à la classe 1 (resp. classe 2), c'est-à-dire s'il n'y a pas de conflit entre les deux classes, nous avons alors  $A_1(x) = 1$  et  $A_2(x) = 0$  (resp.  $A_1(x) = 0$  et  $A_2(x) = 1$ ). Dès lors, nous obtenons  $\Lambda(x) = \mu_1$  (resp.  $\Lambda(x) = \mu_2$ ). Ainsi, la moyenne conditionnelle,  $\Lambda(x)$ , est constante sur les sous-ensembles de niveaux de gris pour lesquels il n'y a pas de conflit entre les classes (voir figure 1.b). La variance  $\Sigma^2(x)$  de la moyenne des voisins de  $s$ ,  $\mu_{vois}(s)$ , est quant à elle donnée par :

$$\Sigma^2(x) = \frac{1}{N(x)} \sum_{s: \omega_s = x} \mu_{vois}(s)^2 - (A_1(x) \mu_1 + A_2(x) \mu_2)^2, \quad (11)$$

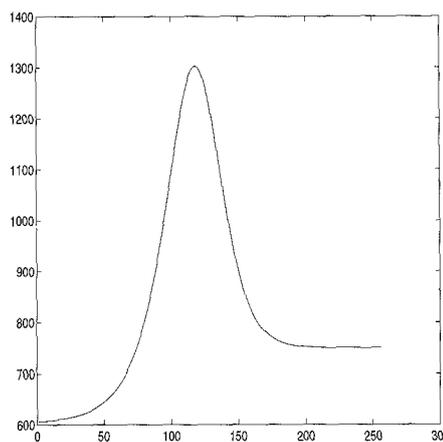
où  $N(x)$  est le nombre de pixels dont le niveau de gris est égal à  $x$ .



a) Somme de deux gaussiennes  $(\mu_1, \sigma_1) = (100, 600)$   $(\mu_2, \sigma_2) = (150, 750)$ .



b) Moyenne conditionnelle  $\Lambda(x)$ .



c) Variance conditionnelle  $\Sigma^2(x)$ .

Figure 1. – Détection d'un mélange par les moments conditionnels.

Alors :

$$\begin{aligned} \Sigma^2(x) &= \frac{1}{N(x)} \left( \sum_{s: \omega_s = x, s \in C_1} \mu_{vois}(s)^2 + \sum_{s: \omega_s = x, s \in C_2} \mu_{vois}(s)^2 \right) \\ &\quad - (A_1(x) \mu_1 + A_2(x) \mu_2)^2. \end{aligned} \quad (12)$$

Si nous ajoutons l'hypothèse classique concernant l'indépendance des pixels conditionnellement à leur classe, nous obtenons :

$$\Sigma^2(x) = A_1(x)(\sigma_1^2 + \mu_1^2) + A_2(x)(\sigma_2^2 + \mu_2^2) - (A_1(x)\mu_1 + A_2(x)\mu_2)^2, \tag{13}$$

c'est-à-dire :

$$\Sigma^2(x) = A_1(x)\sigma_1^2 + A_2(x)\sigma_2^2 + A_1(x)(1 - A_1(x))\mu_1^2 + A_2(x)(1 - A_2(x))\mu_2^2 - 2A_1(x)A_2(x)\mu_1\mu_2. \tag{14}$$

Comme nous avons  $A_1(x) + A_2(x) = 1$ , nous obtenons :

$$\Sigma^2(x) = A_1(x)\sigma_1^2 + A_2(x)\sigma_2^2 + A_1(x)A_2(x)(\mu_1 - \mu_2)^2. \tag{15}$$

En conclusion, le même argument que pour la moyenne conditionnelle permet d'affirmer que la variance est constante pour les zones de l'espace des niveaux de gris où il n'y a pas de conflit entre classes (voir figure 1.c). Posons  $\alpha = A_1(x)$ . Alors  $A_2(x) = 1 - \alpha$  et entre deux classes successives, le long de la zone de conflit, nous avons :

$$\Sigma^2(\alpha) = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 + \alpha(1 - \alpha)(\mu_1 - \mu_2)^2, \alpha \in ]0, 1[. \tag{16}$$

Cette fonction, dépendant de la variable  $\alpha$ , est une parabole. Nous avons  $\Sigma^2(0) = \sigma_2^2$  et  $\Sigma^2(1) = \sigma_1^2$ . Il est alors aisé de montrer que, si  $|\sigma_1^2 - \sigma_2^2| < (\mu_1 - \mu_2)^2$ , alors il existe  $\alpha \in ]0, 1[$  tel que,  $\Sigma^2(\alpha) > \Sigma^2(0)$  et  $\Sigma^2(\alpha) > \Sigma^2(1)$ . Si ce maximum local est compris entre  $\mu_1$  et  $\mu_2$ , il permet de détecter la zone de conflit entre les deux classes (voir figure 1.c).

## 2.2. détection des modes par une analyse dans un espace d'échelle

Nous présentons ici une méthode qui est notamment utilisée pour mettre en correspondance une courbe avec une somme de gaussiennes [15]. Un mode sur une courbe est défini comme un maximum local. Nous proposons d'utiliser une analyse dans un espace d'échelle, connue sous l'acronyme anglais SSA ("Scale Space Analysis") pour détecter les modes de l'histogramme de la variance conditionnelle (HVC), qui représentent les zones de conflits entre classes. Cette méthode, proposée par Witkin [22] a été appliquée pour détecter les modes de l'histogramme d'une image dans [3]. Considérons une séquence de données monovaluées et uniformément espacées, par exemple un histogramme. Considérons alors la convolution de ce signal par un noyau gaussien dont la variance varie dans un intervalle donné. Chaque convolution est alors associée à une résolution. A chaque résolution, nous calculons les points de passage à zéro de la dérivée seconde, qui correspondent aux points d'inflexion. Sur l'histogramme d'une image, deux points d'inflexions successifs (passage négatif-positif de la dérivée seconde puis passage positif-négatif) permettent de définir une classe. Les intervalles entre les classes définissant les zones de conflit. Cette méthode est utilisée dans [12] pour définir automatiquement le terme d'attache aux données d'une modélisation markovienne, utilisée pour segmenter l'ALD (Adréno-leukodystrophie) à partir d'échos IRM (Images à Résonance Magnétique).

Pour illustrer notre propos, considérons une distribution gaussienne, caractérisée par sa moyenne  $\mu$  et sa variance  $\sigma^2$  (voir

figure 2.a). Si nous convoluons cette distribution par un filtre gaussien centré, caractérisé par sa variance  $\sigma_f^2$ , nous obtenons une gaussienne, dont la moyenne est donnée par  $\mu$  et la variance par  $\sqrt{\sigma^2 + \sigma_f^2}$ . Ainsi, lorsque la variance du filtre varie, le lieu des points  $y$  où la dérivée seconde s'annule est défini de la façon suivante :

$$y = \mu \pm \sqrt{\sigma^2 + \sigma_f^2}. \tag{17}$$

Dans ce cas, l'espace d'échelle est donc constitué de deux arcs de courbes divergents (voir figure 2.b).

Considérons maintenant une distribution définie par la somme de deux gaussiennes (voir figure 2.c). La convolution par un noyau gaussien reste la somme de deux gaussiennes. Si les deux gaussiennes sont suffisamment séparées, nous obtenons une courbe bimodale et la dérivée seconde possède quatre zéros. Dans le cas contraire, la dérivée seconde ne présente que deux zéros. En haut de l'espace d'échelle (pour de faibles variances du noyau gaussien), nous avons donc quatre points correspondant aux changements de signe de la dérivée première. Lorsque  $\sigma_f^2$  croît, les deux zéros centraux se rapprochent et finissent par se confondre pour disparaître (voir figure 2.d). L'espace d'échelle de la somme de deux gaussiennes consiste donc en deux courbes divergentes et une arche entre les deux. Notons que si les deux gaussiennes initiales sont fortement mélangées, nous obtenons uniquement les deux courbes divergentes.

Cette SSA peut donc être utilisée pour estimer le nombre de modes dans un mélange. De plus, la localisation des arches permet d'obtenir une première estimation des paramètres des distributions associées aux modes et peut donc être utilisée comme initialisation pour un algorithme de mise en correspondance avec une somme de gaussiennes, tel que l'algorithme de Marquart [10].

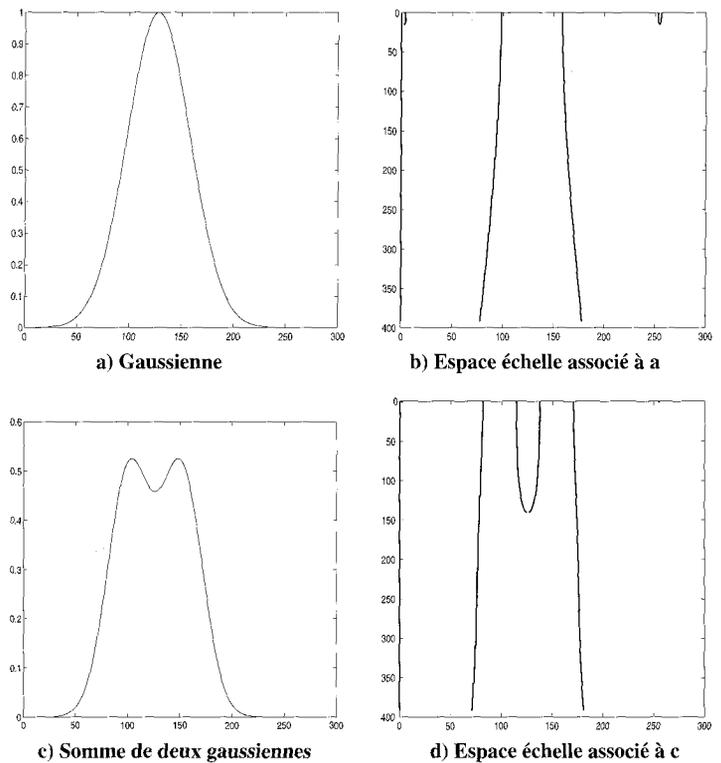


Figure 2. - Gaussiennes et espaces échelle associés.

Nous proposons d'utiliser cette technique pour détecter les différents modes d'un HVC. Nous n'avons pas besoin d'une hypothèse gaussienne mais supposons simplement que chaque mode est caractérisé par un point d'inflexion de chaque côté. L'utilisation d'une SSA permet d'éviter de mauvaises détections dues au bruit. Considérons l'espace d'échelle associé avec un HVC donné. A partir de la convolution du HVC par un noyau gaussien de variance  $\sigma_f^2$ , nous calculons l'ensemble des points où la dérivée seconde s'annule. A partir de cet ensemble, nous définissons les sous-ensembles de niveaux de gris correspondant aux classes et aux zones de conflit entre classes.

L'algorithme se définit alors de la façon suivante : à partir des données, nous estimons d'abord la variance  $\Sigma^2(x)$  de la moyenne des voisins de  $s$  comme une fonction de  $x$ , où  $x$  représente le niveau de gris du pixel  $s$ . En pratique, le HVC est constitué de paliers plus ou moins constants, correspondant aux différentes classes, séparés par des modes de valeurs plus élevées, correspondant aux zones de conflit. Nous détectons alors les modes du HVC en utilisant une SSA. En pratique, nous convoluons le signal par un unique noyau gaussien. La variance de ce noyau est un paramètre donné par l'utilisateur. Plus cette variance est grande, moins nous détectons de classes. La valeur de ce paramètre dépend donc de la finesse de la classification désirée. Pour finir, nous estimons les passages par zéros de la dérivée seconde du résultat de la convolution. A partir de deux zéros définissant un mode, nous définissons une zone de conflit, ce qui permet ensuite de définir les différentes classes et leur fonction d'appartenance, comme indiqué sur la figure 3.

Considérons un cas particulier pour expliciter la méthode. Soit un HVC correspondant à quatre classes (voir figure 3.a). L'espace d'échelle associé et la définition des classes et des zones de conflit sont montrés respectivement sur les figures 3.b et 3.c. Sur le sous-ensemble noté B sur la figure 3.c, nous considérons que les niveaux de gris correspondent à la classe B. Sur les sous-ensembles notés A & B et B & C, nous définissons une zone de conflit entre la classe B et respectivement les classes A et C. Nous obtenons alors une classification floue, interpolant linéairement les fonctions d'appartenance sur les zones A & B et B & C.

### 3. domaine de validité dans le cas d'un mélange de deux gaussiennes

En pratique, un mode correspondant à une zone de conflit ne sera valide que si le maximum local correspondant se situe entre les moyennes associées aux deux distributions. Le domaine de validité de la méthode est défini par ce critère. Nous comparons ici ce domaine avec le domaine correspondant à la bimodalité dans le cas d'un mélange de deux gaussiennes. Nous pourrions ainsi démontrer l'apport de l'hypothèse contextuelle. Considérons un tel mélange défini par la probabilité *a priori* de la première classe  $\pi$ , les moyennes  $\mu_1$  et  $\mu_2$ , et les écarts type  $\sigma_1$  and  $\sigma_2$ . La loi du mélange s'écrit alors de la façon suivante :

$$p(x) = \pi p_1(x) + (1 - \pi)p_2(x), \quad (18)$$

où  $p_1 = \mathcal{N}(\mu_1, \sigma_1^2)$  et  $p_2 = \mathcal{N}(\mu_2, \sigma_2^2)$ ;  $\mathcal{N}(a, b)$  représentant une loi normale de moyenne  $a$  et de variance  $b$ . Considérons la fonction  $\Sigma^2(\alpha)$  définie précédemment. Le maximum local associé au mélange considéré est défini par l'équation suivante :

$$\frac{d\Sigma^2(\alpha)}{d\alpha} = -2\alpha(\mu_1 - \mu_2)^2 + \sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2 = 0. \quad (19)$$

Le maximum est alors localisé par l'équation :

$$\text{Arg max } \Sigma^2(\alpha) = \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2}. \quad (20)$$

La fonction  $\Sigma^2(\alpha)$  possède un mode entre  $\mu_1$  et  $\mu_2$  si et seulement si :

$$\begin{cases} \alpha(\mu_1) < \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \\ \alpha(\mu_2) > \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \end{cases} \text{ ou } \begin{cases} \alpha(\mu_1) > \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \\ \alpha(\mu_2) < \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \end{cases} \quad (21)$$

En reprenant les notations de [14], nous définissons :

$$\Delta = \frac{\mu_2 - \mu_1}{\sigma_1} \quad \text{and} \quad \sigma = \frac{\sigma_2}{\sigma_1}. \quad (22)$$

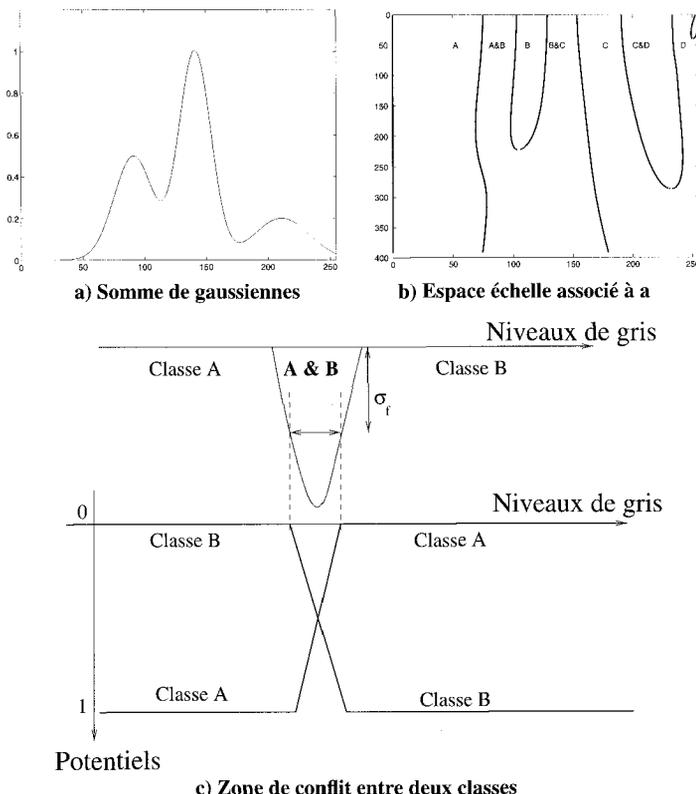


Figure 3. – Définition des classes et des zones de conflit.

Sans perte de généralité, nous prenons  $\Delta \geq 0$  et  $\sigma \leq 1$ . La condition nécessaire et suffisante définie par (21) peut alors se résumer par les inégalités suivantes :

$$\begin{cases} \frac{1}{1 + (\frac{1}{\pi} - 1)\frac{1}{\sigma} \exp - \frac{\Delta^2}{2\sigma^2}} > \frac{1 - \sigma^2 + \Delta^2}{2\Delta^2} \\ \frac{1}{1 + (\frac{1}{\pi} - 1)\frac{1}{\sigma} \exp + \frac{\Delta^2}{2}} < \frac{1 - \sigma^2 + \Delta^2}{2\Delta^2} \end{cases} \quad (23)$$

En utilisant les notations suivantes :

$$E_\sigma = \frac{1}{\sigma} \exp - \frac{\Delta^2}{2\sigma^2} \quad \text{and} \quad E_1 = \frac{1}{\sigma} \exp + \frac{\Delta^2}{2}, \quad (24)$$

le domaine de validité est alors défini par :

$$\begin{cases} \frac{1}{\pi} < \frac{1}{E_\sigma} \frac{\Delta^2 - 1 + \sigma^2}{\Delta^2 + 1 - \sigma^2} + 1, \\ \frac{1}{\pi} > \frac{1}{E_1} \frac{\Delta^2 - 1 + \sigma^2}{\Delta^2 + 1 - \sigma^2} + 1. \end{cases} \quad (25)$$

Une condition nécessaire et suffisante de bimodalité dans le cas d'un mélange de deux gaussiennes peut être trouvée dans [14]. Cette condition s'exprime de la façon suivante. Considérons  $\Delta_0$  donné par :

$$\Delta_0 = \frac{1}{\sigma} [2(\sigma^4 - \sigma^2 + 1)^{3/2} - (2\sigma^6 - 3\sigma^4 - 3\sigma^2 + 2)]^{1/2}. \quad (26)$$

Le mélange est alors bimodal si et seulement si  $\Delta > \Delta_0$  et  $\pi$  se trouve dans l'intervalle ouvert  $]\pi_1, \pi_2[$ , avec :

$$\pi_i^{-1} = 1 + \frac{\sigma^3 y_i}{\Delta - y_i} \exp \left( -\frac{1}{2} y_i^2 + \frac{1}{2} \frac{(y_i - \Delta)^2}{\sigma^2} \right), \quad i = 1, 2, \quad (27)$$

où  $(y_1, y_2)$  sont les racines de :

$$(\sigma^2 - 1)y^3 - \Delta(\sigma^2 - 2)y^2 - \Delta^2 y + \Delta\sigma^2 = 0, \quad (28)$$

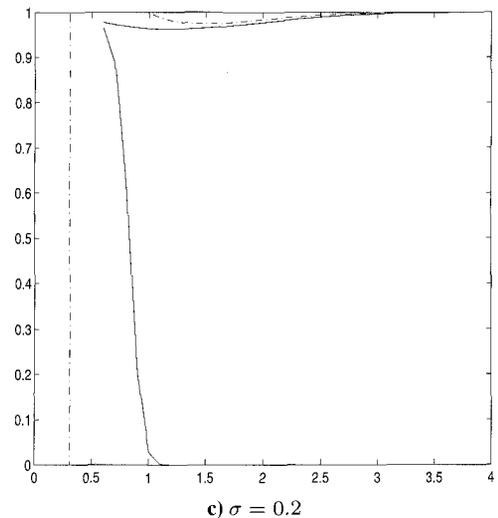
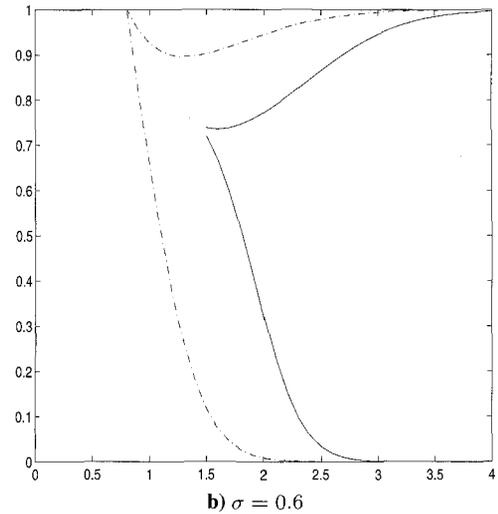
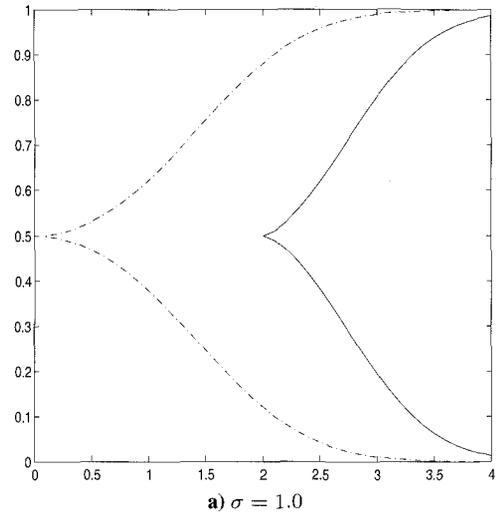
avec  $0 < y_1 < y_2 < \Delta$ .

La figure 4 montre les domaines respectifs où le mélange peut être détecté avec une SSA sur l'histogramme et sur le HVC. Le domaine de validité des deux approches est représenté par les zones se trouvant à droite des courbes. La bimodalité de l'histogramme est définie par les courbes pleines alors que le domaine de validité de la méthode proposée est défini par les courbes en pointillés. Ainsi, la prise en compte d'une information contextuelle permet d'étendre considérablement le domaine de validité.

## 4. segmentation par une estimation itérative conditionnelle

### 4.1. principe des algorithmes de type EM

Pour étudier les performances de la méthode de classification proposée, nous l'insérons dans un algorithme de type EM



Axe des X :  $\Delta = \frac{\mu_2 - \mu_1}{\sigma_1}$ , Axe des Y :  $\pi = \frac{P(s \in C_1)}{P(s \in C_2)}$ .

Courbes pleines : bimodalité, courbes pointillées : méthode HVC. Figure 4. - Détection d'un mélange par les moments conditionnels.

("expectation-maximisation") [20]. Ce type d'algorithmes a été introduit en traitement des images à la fin des années 80 [4], [5] et a fait l'objet de plusieurs variantes comme le SEM [16] ou l'ICE [17], [18]. Le principe de ces algorithmes est d'alterner des séquences d'optimisation du modèle et d'estimation de ses paramètres. Ainsi, une initialisation de l'estimation est affinée après chaque étape d'optimisation, qui prend en compte l'information de type contextuel par l'intermédiaire du terme de régularisation. Nous appliquons ici le principe de l'estimation itérative conditionnelle (ICE) :

1. Initialisation des paramètres de l'attache aux données par la méthode HVC par les estimateurs empiriques des moyennes et variances,
2. Générer  $n$  réalisations du champ des labels par un échantillonneur de Gibbs avec l'estimée courante des paramètres,
3. Actualiser les paramètres des classes conditionnellement au champ des labels,
4. Retourner à l'étape 2 jusqu'à ce que les paramètres associés aux classes soient stabilisés.

## 4.2. résultats

Nous considérons en premier lieu une image de synthèse pour laquelle nous connaissons donc le nombre de classes ainsi que leur moyenne et leur variance (voir figure 5.a). Cette image est bruitée par un bruit additif gaussien (voir figure 5.b). Une seconde image permet de valider ce travail sur des données réelles. Il s'agit d'une image SPOT (canal XS3) de la Hollande (voir figure 6). La scène représente une zone de parcellaire, possédant des structures très fines (routes, chemins, champs étroits,...).

Les résultats sont obtenus pour deux modèles *a priori* différents : le modèle de Potts et un modèle préservant les structures fines (le Chien-modèle). Une description et une étude comparative de ces deux modèles se trouvent dans [23]. La table 1 montre les valeurs des paramètres estimés pour chaque étape de l'ICE avec respectivement le modèle de Potts et le Chien-modèle comme loi *a priori*. Les résultats obtenus sur les images de synthèse prouvent les bonnes performances de la méthode de l'HVC. Les résultats des segmentations aboutissent aux mêmes conclusions que dans [23] concernant les modèles *a priori*. Notons que l'apport du principe EM est le même pour les deux modèles (voir figure 7). Néanmoins, le chien-modèle, étant plus fin dans

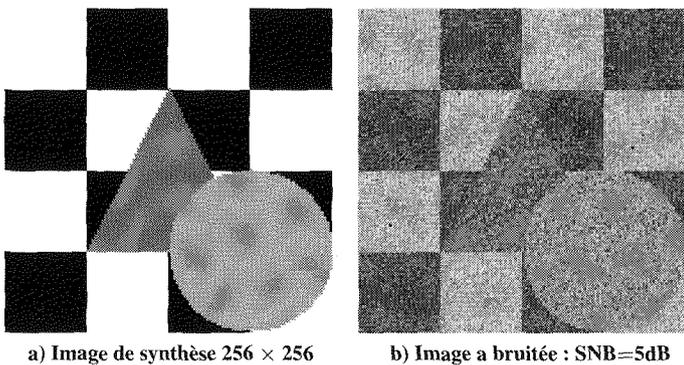


Figure 5. – Image test de synthèse.

Tableau 1. – Estimation des paramètres associés aux classes de l'image de synthèse #1 lors d'un ICE initialisé par un HVC pour les modèle de Potts (P.) et chien-modèle (C.). Les valeurs réelles sont portées dans la colonne V.R.

	HVC	P. (it. 1)	C. (it. 1)	P. (it. 2)	C. (it. 2)	P. (it. 3)	C. (it. 3)	V. R.
$\mu_1$	70	84	78	84	82	84	84	85
$\sigma_1^2$	333	477	386	497	459	498	485	500
$\mu_2$	107	113	107	114	111	114	113	115
$\sigma_2^2$	233	439	312	487	382	489	437	500
$\mu_3$	147	145	145	144	144	144	144	145
$\sigma_3^2$	300	522	418	550	500	553	520	500
$\mu_4$	185	175	177	174	175	174	175	175
$\sigma_4^2$	400	497	447	504	491	504	498	500



Figure 6. – Image SPOT (canal XS3).

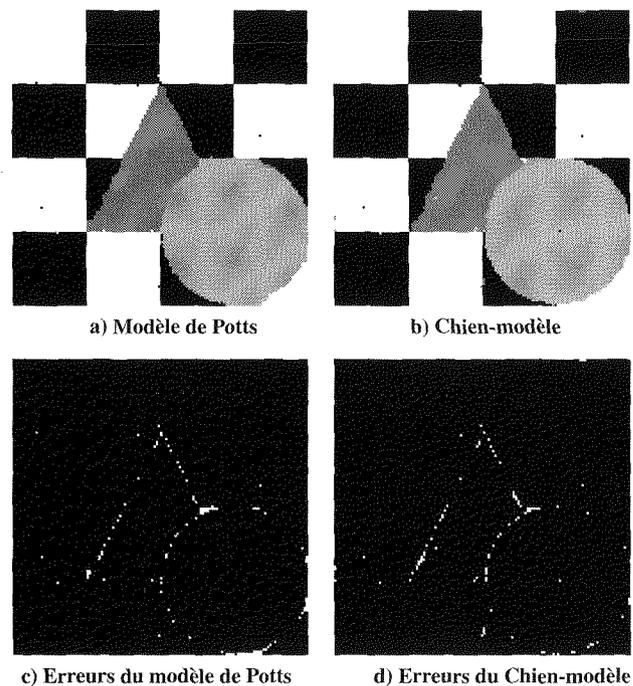


Figure 7. – Résultats obtenus pour l'image de synthèse #1 avec un ICE initialisé par un HVC.

la régularisation, nécessite un plus grand nombre d'itérations pour arriver à la convergence des paramètres de l'attache aux données. Pour le modèle de Potts, une itération est suffisante pour la convergence des estimées, ce qui revient à utiliser deux itérations d'optimisation. En revanche, le chien-modèle nécessite quatre séquences d'optimisation. Les figures 8 et 9 montrent les résultats obtenus avec respectivement le modèle de Potts et le chien-modèle. Les structures fines sont nettement mieux préservées avec le chien-modèle.



Figure 8. – Résultats obtenus pour l'image SPOT avec un ICE initialisé par un HVC et un modèle de Potts.



Figure 9. – Résultats obtenus pour l'image SPOT avec un ICE initialisé par un HVC et un chien-modèle.

## 5. à propos du choix du modèle d'attache aux données

Dans le paragraphe précédent, nous avons proposé une méthode pour situer les différentes classes sur l'histogramme. Il reste désormais à définir le modèle d'attache aux données à partir de cette estimation pour l'incorporer dans un cadre markovien.

### 5.1. entre loi bayésienne et relaxation d'étiquettes

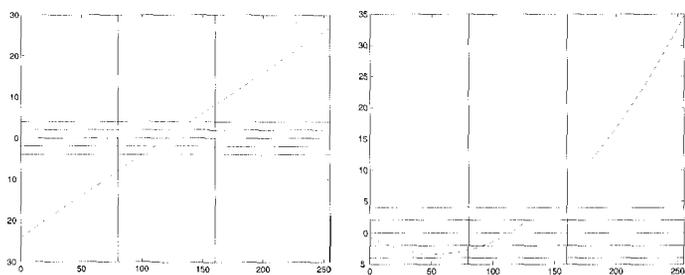
Nous présentons dans ce paragraphe deux approches pour définir le terme d'attache aux données. En premier lieu, considérons une segmentation fondée rigoureusement sur la loi de Bayes. Nous devons alors optimiser un critère bayésien fondé sur le produit  $P(Y|X)P(X)$ . Avec une telle approche, le terme d'attache aux données pour une modélisation markovienne est donc défini par la fonction :  $-\log P(X)$ . Définir le terme d'attache aux données revient alors à modéliser la texture (ou le bruit) pour chacune des classes [6], [8]. Supposons pour simplifier que le bruit sur chaque classe est gaussien et que les pixels sont indépendants conditionnellement à leur classe. Les paramètres de chaque classe (moyenne et variance) sont estimés à partir du HVC. La loi de Bayes impose alors de définir le terme d'attache aux données de la façon suivante :

$$U_{att}(X) = \sum_{s \in S} \sum_{\lambda \in \Lambda} \frac{(x_s - \mu_\lambda)^2}{\sigma_\lambda^2} + \frac{1}{2} \log 2\pi\sigma_\lambda^2 \delta_{y_s=\lambda}. \quad (29)$$

Une approche, plus proche de la physique, consiste à faire une première classification ponctuelle à partir de l'estimation des classes, par Maximum de Vraisemblance par exemple. Une relaxation des étiquettes est ensuite effectuée en utilisant le modèle *a priori*. Le terme d'attache aux données est alors défini à partir de la première classification, considérée comme un champ externe hétérogène [19]. Nous notons  $Y^0 = (y_s^0)_{s \in S}$  le résultat de la classification ponctuelle. Le terme d'attache aux données lors de la relaxation est alors défini par :

$$U_{att}(X) = \alpha \sum_{s \in S} \delta_{y_s=y_s^0}. \quad (30)$$

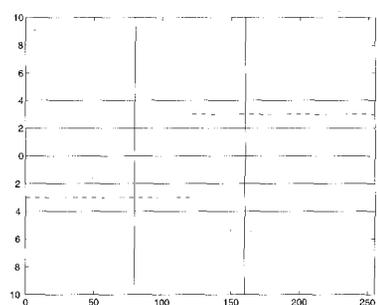
Pour étudier plus précisément le comportement de l'énergie associée à ces deux approches, nous considérons le cas d'un mélange de deux classes gaussiennes, en utilisant un modèle d'Ising 4-connexe. Ces restrictions permettent de simplifier l'étude, mais les conclusions s'étendent aisément à un plus grand nombre de classes, de lois différentes et pour d'autres modèles de régularisation. Nous notons  $U_1$  (resp.  $U_2$ ) le terme d'attache aux données relativement à la classe 1 (resp. 2) définie par la loi normale  $\mathcal{N}(\mu_1, \sigma_1^2)$  (resp.  $\mathcal{N}(\mu_2, \sigma_2^2)$ ).  $U_R$  représente l'énergie due au terme de régularisation. Nous supposons sans perte de généralité que  $\mu_1 < \mu_2$ . La figure 10 montre les différents types d'attache aux données. Si nous considérons l'énergie induite par les lois gaussiennes, le comportement de l'énergie entre les moyennes  $\mu_1$  et  $\mu_2$  correspond à ce que nous pouvons attendre. En effet, la différence entre les énergies d'attache aux données va croître entre  $\mu_1$  et  $\mu_2$ . De plus, cette croissance va dépendre des variances et donc prendre en compte la connaissance que nous avons des données. En revanche, en dehors de cet intervalle, la différence des énergies va tendre vers l'infini. Or, l'énergie liée au terme de régularisation est quantifiée et reste finie (courbes horizontales pleines sur la figure 10). Par conséquent, si la valeur d'une donnée est extrême,



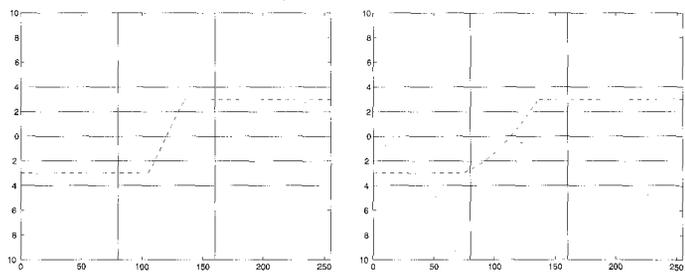
$\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 20,$

$\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 40$

a) Attache aux données définie par la loi des classes



b) Attache aux données pour une relaxation



$\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 20,$

$\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 40$

c) Attache aux données proposée

Courbes en pointillés : différence des énergies d'attache aux données  $U_1 - U_2$ .

Courbes pleines : différences des énergies de régularisation (modèle de Potts).

Les droites verticales représentent les moyennes  $\mu_1$  et  $\mu_2$ .

Axe des X : Niveaux de gris.

Figure 10. - Résultats obtenus pour l'image de synthèse # 1 avec un ICE initialisé par un HVC.

le pixel sera classifié dans la classe correspondante indépendamment du contexte. En pratique, le résultat d'une segmentation restera bruité en un tel pixel. Si nous considérons l'énergie induite par une relaxation, nous avons un compromis entre attache aux données et régularisation, quelles que soient les valeurs des données. Néanmoins, la définition en tout ou rien de l'énergie d'attache aux données ne rend pas compte de l'information que nous possédons sur la loi des différentes classes. Nous proposons donc d'utiliser un terme d'attache aux données intermédiaire en privilégiant l'énergie induite par la loi des classes mais en la bor-

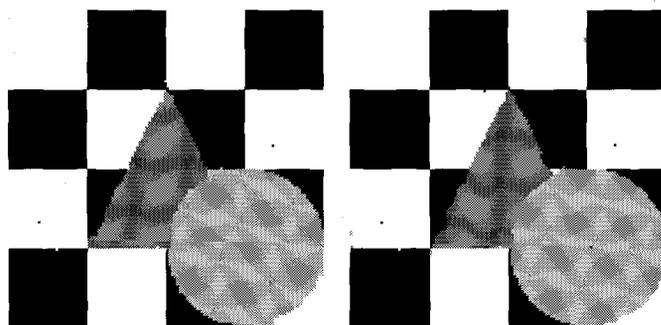
nant de sorte qu'elle reste inférieure à l'énergie maximale du terme de régularisation. Nous considérons alors dans le cadre de la 4-connexité :

$$V_{c=\{s\}}(x_s | y_s = \lambda) = \frac{(x_s - \mu_\lambda)^2}{\sigma_\lambda^2}, \quad \text{si } \frac{(x_s - \mu_\lambda)^2}{\sigma_\lambda^2} < 3\beta,$$

$$V_{c=\{s\}}(x_s | y_s = \lambda) = 3\beta \quad \text{sinon} \quad (31)$$

## 5.2. influence sur les résultats pratiques

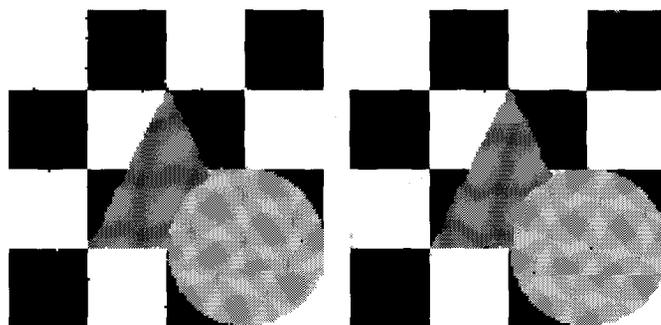
En pratique, le fait de borner le terme d'attache aux données permet de supprimer tous les points isolés dans le résultat final sans pour autant sur-régulariser en donnant trop d'importance au modèle *a priori*. Les résultats de la figure 11 mettent partiellement en valeur ce phénomène. Néanmoins, pour ces images test, le bruit considéré a été ajouté synthétiquement et suit donc réellement une loi gaussienne, ce qui limite la portée des remarques précédentes. En pratique, le modèle du bruit n'est qu'une approximation et certains pixels peuvent avoir un niveau de gris très éloigné du centre de la classe à laquelle ils appartiennent. Dans ce cas, il est nécessaire de borner le terme d'attache aux données pour éviter une mauvaise classification de ces points, due à une énergie d'attache aux données trop forte.



a) Modèle de Potts :  $\#E = 104$

b) Chien-modèle :  $\#E = 93$

Attache aux données quadratique ( $\#E$  : nombre d'erreurs)



c) Modèle de Potts :  $\#E = 99$

d) Chien-modèle :  $\#E = 85$

Attache aux données bornée ( $\#E$  : nombre d'erreurs)

Figure 11. - Influence du terme d'attache aux données.

## 6. conclusion

Dans cet article, nous avons abordé le problème de la classification : estimation du nombre de classes et des paramètres associés à ces classes. Ces estimées permettent de définir le terme d'attache aux données d'une modélisation markovienne. Nous proposons une méthode utilisant une hypothèse contextuelle. L'apport de cette hypothèse est théoriquement étudié dans le cas d'un mélange de deux gaussiennes (image contenant deux classes de lois gaussiennes). La méthode est ensuite validée sur des données synthétiques et réelles et intégrée dans un algorithme non-supervisé de type ICE. Pour être complet, ce travail devrait aborder l'estimation des paramètres du modèle *a priori*. Certaines méthodes sont proposées dans la littérature [2], [9]. Nous étudions actuellement une méthode fondée sur une chaîne de Markov dont la convergence est obtenue par une méthode de Monte Carlo (MCMC). Cette méthode permet notamment d'estimer la fonction de partition, ce qui induit des estimateurs au sens du Maximum de Vraisemblance (MCMCML) [21]. Ce dernier point permettra d'obtenir une méthode totalement non-supervisée de segmentation.

Pour finir, nous avons comparé et analysé différents types d'attache aux données. Les résultats obtenus nous font préférer une attache aux données intermédiaire entre celle induite par la loi des classes et celle issue de la relaxation.

### BIBLIOGRAPHIE

- [1] J. Besag. Spatial Interaction and Statistical Analysis of Lattice Systems. *J. Royal Statistical Society, Series B*, 36 : 721-741, 1974.
- [2] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statistical Society, Series B*, 48 : 259-302, 1986.
- [3] M.J. Carlotto. Histogram analysis using a Scale-Space approach. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 9(1) : 121-129, January 1987.
- [4] B. Chalmoud. Image restoration using an estimated Markov model. *Signal Processing*, 15 : 115-129, 1988.
- [5] J. Zerubia, R. Chellappa. Mean field approximation using Compound Gauss Markov Random fields for edge detection and image estimation. *IEEE trans. on Neural Networks*, 8(4) : 703-709, July 1993.
- [6] C.S. Won, H. Derin. Unsupervised segmentation of noisy and textured images using Markov Random Fields. *Computer Vision Graphics and Image Processing*, 4 : 308-328, 1992.
- [7] S. Lakshmanan, H. Derin. Simultaneous Parameter Estimation and Segmentation of Gibbs Random Fields Using Simulated *IEEE trans. on Pattern Analysis and Machine Intelligence*, 11(8) : 799-813, August 1989.
- [8] X. Descombes. A fission and fusion Markovian approach for multi-channel segmentation. In *Proc. IGARSS'95, Firenze, Italy*, pages 124-127, 1995.
- [9] H. Derin, H. Elliott. Modelling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 9 (1) : 39-55, January 1987.
- [10] W. Press, S. Teukolski, W. Vetterling, B. Flannery. *Numerical Recipes in C : The Art of Scientific Computing*. Cambridge University Press, 2nd ed., 1992.
- [11] S. Geman, D. Geman. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 6 (6) : 721-741, 1984.
- [12] L. Aurdal, X. Descombes, H. Maître, I. Bloch, C. Adamsbaum, G. Kalifa. Fully Automated Analysis of adrenoleukodystrophy from dual echo MR-images. In *CAR'95 Berlin*, pages 35-40, 1995.
- [13] Duda, P.E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York, 1973.
- [14] D.M. Titterton, A.F.M. Smith, U.E. Makov. *Statistical Analysis of finite mixture Distributions*. John Wiley and Sons, 1985.
- [15] A. Goshtasby, W.D. O'Neill. Curve fitting by a sum of Gaussians. *CVGIP : Graphical Models and Image Processing*, 56(4) : 281-288, 1994.
- [16] P. Masson, W. Pieczynski, SEM Algorithm and Unsupervised Statistical Segmentation of Satellite Images. *IEEE trans. on Geoscience and Remote Sensing* 31(3) : 618-633, May 1993.
- [17] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative; *Traitement du signal*, 11(2), 1994.
- [18] Z. Kato, M. Berthod, J. Zerubia, W. Pieczynski, Unsupervised adaptive image segmentation. In *proc. ICASSP'95, Detroit, Michigan, USA*, pages 2399-2402, May 1995.
- [19] M. Sigelle, R. Ronfard. Relaxation of previously classified images by a Markov field technique and its relationship with statistical physics. In *Proc. 7th SCIA 91, Aalborg, Danmark*, pages 387-394, 1991.
- [20] A.P. Dempster, N.M. Laird, D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *J. Royal Statistical Society, serie B*, 39 (1) : 1-38, 1977.
- [21] C. J. Geyer, E.E. Thompson. Constrained Monte Carlo Maximum Likelihood for Dependent Data. *J. Royal Statistical Society, serie B*, 54(3) : 657-699, 1992.
- [22] A.P. Witkin, Scale space filtering : a new approach to multi-scale description. *Proc. Image Understanding*, pages 79-95, 1984.
- [23] X. Descombes, R. Morris, J. Zerubia. Quelques améliorations à la segmentation d'images bayésienne. Première partie : modélisation. *traitement du signal*, 1997.

Manuscrit reçu le 18 juillet 1996

## LES AUTEURS

### Xavier DESCOMBES



Xavier Descombes est chercheur invité à l'institut Max Planck pour les Sciences Cognitives à Leipzig (Allemagne) depuis octobre 1996. Auparavant il a travaillé en tant que chercheur post-doctorant successivement à l'ENST (1994), à l'Université Catholique de Louvain (1995) et à l'INRIA au sein du projet PASTIS (1996). Il a obtenu un diplôme d'ingénieur de l'ENST en 89, un DEA de mathématiques pures à l'université Pierre et Marie Curie (Paris VI) en 90 et une thèse de doctorat à l'ENST en 93. Ses domaines d'intérêt sont la modélisation stochastique, l'estimation des paramètres et l'optimisation. Ses travaux, concernant principalement les champs de Markov, ont été appliqués à l'imagerie satellitaire et médicale pour des tâches de segmentation, de restauration et d'extraction d'objets.

### Robin MORRIS



Robin Morris a obtenu le "BA degree" au département ingénierie de l'université de Cambridge en 1991. Il obtient son doctorat dans le groupe de traitement du signal du même département en 1995. En Octobre 1994, il obtient une bourse "Junior Research Fellowship" du Trinity College de Cambridge. En 1996, il est à l'INRIA. Depuis Janvier 1997, il travaille dans le groupe d'analyse de données du "NASA Ames research center" grâce à une bourse NRC. Ses domaines d'intérêt sont l'inférence Bayésienne, la modélisation d'images et les méthodes MCMC.

### Josiane ZÉRUBIA



Josiane Zérubia est chercheur permanent à l'INRIA depuis 1989. Elle est directeur de recherche depuis juillet 1995 et responsable du laboratoire PASTIS (INRIA Sophia-Antipolis) depuis novembre 1995. Auparavant, elle était à l'institut de traitement des signaux et des images de l'université de Californie du sud à Los Angeles en tant que Post doctorante. Elle a également travaillé comme chercheur au LASSY (Université de Nice et CNRS) entre 1984 et 1988, au laboratoire de recherche de Hewlet Packard en France et à Palo-Alto entre 82 et 84. Elle a obtenu un diplôme d'ingénieur de l'ENSIEG, Grenoble, en 81, un doctorat ingénieur en 86, un doctorat en 88 et une habilitation à diriger des recherches en 94. Elle est membre de l'IEEE depuis 1988 et de l'académie des sciences de New-York depuis 1996. Son domaine d'intérêt est le traitement des images, en particulier la restauration d'images, la segmentation, classification, le groupement perceptuel, la super-résolution en utilisant des modèles probabilistes ou des réseaux neuronaux. Elle travaille également sur l'estimation de paramètres et les techniques d'optimisation.