



HAL
open science

Quelques améliorations à la segmentation d'images bayésienne

Xavier Descombes, Robin Morris, Josiane Zerubia

► **To cite this version:**

Xavier Descombes, Robin Morris, Josiane Zerubia. Quelques améliorations à la segmentation d'images bayésienne. RR-2916, INRIA. 1996. inria-00073781

HAL Id: inria-00073781

<https://inria.hal.science/inria-00073781>

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*Quelques améliorations à la segmentation
d'images bayésienne*

Xavier Descombes, Robin Morris, Josiane Zerubia

N° 2916

Juin 1996

————— THÈME 3 —————



*Rapport
de recherche*

Quelques améliorations à la segmentation d'images bayésienne

Xavier Descombes, Robin Morris, Josiane Zerubia

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet PASTIS

Rapport de recherche n° 2916 — Juin 1996 — 55 pages

Résumé : Nous détaillons dans ce rapport les différentes étapes de la segmentation Bayésienne : modélisation, estimation, optimisation. La modélisation est appréhendée sous l'angle des Champs de Markov. Nous montrons les limites du modèle de Potts couramment employé et proposons un nouveau modèle (le chien-modèle) permettant de contrôler la longueur des contours et des lignes dans l'image segmentée. Nous préservons ainsi les structures fines présentes dans les données. Nous abordons ensuite le problème de l'estimation du nombre de classes et des paramètres qui leur sont associés. Nous proposons une méthode utilisant l'hypothèse contextuelle inhérente aux images pour discriminer les différentes classes dans le cas de distributions fortement mélangées. Pour finir, nous comparons les critères MPM et MAP conjointement aux algorithmes qui permettent de les optimiser. Les différents résultats sont obtenus sur des images synthétiques et des images SPOT.

Mots-clé : Segmentation d'images, Champs de Markov, Estimation, Mélange de distributions, Critères Bayésiens

(Abstract: pto)

* . Les auteurs tiennent à remercier le GdR TdSI pour l'image de synthèse n°2 ainsi que le CNES pour l'image SPOT

** . email: *name@sophia.inria.fr*

Some Improvements to Bayesian Image Segmentation

Abstract: We detail in this report the different stages involved in Bayesian image segmentation: modelling, estimation and optimisation. We consider modelling by Markov random fields. We demonstrate the limitations of the Potts model currently employed, and propose a new model (the chien model) which allows us to control the boundary length and lines in the segmented images. We also preserve fine structures in the data. Next 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 between the different classes in the case of mixture distributions with strongly mixed classes. Finally, we compare the MPM (maximiser of the posterior marginals) and MAP (maximum a-posteriori) criteria when used with the algorithms discussed above. Results are presented on synthetic images and SPOT imagery.

Key-words: Image segmentation, Markov Random Fields, Estimation, Mixtures, Bayesian framework

Table des matières

1	Introduction	4
1.1	La segmentation Bayésienne avec une loi a priori Markovienne .	4
1.2	Présentation des images	5
2	Un modèle a priori adapté aux images	8
2.1	Limites du modèle de Potts	8
2.2	Le Chien-modèle	14
2.2.1	Construction du modèle binaire	14
2.2.2	Extension à un modèle m-aire	20
2.3	Résultat obtenus par les deux modèles	21
3	Estimation du modèle d’attache aux données	28
3.1	Méthode de l’Histogramme de la Variance Conditionnelle (HVC)	28
3.1.1	Principe	29
3.1.2	Détection des modes par une analyse d’un espace d’échelle	31
3.1.3	Domaine de validité dans le cas d’un mélange de deux gaussiennes	35
3.2	Principe des algorithmes de type EM et résultats	38
4	A propos du choix du modèle d’attache aux données	45
4.1	Entre loi bayésienne et relaxation de labels	45
4.2	Influence sur les résultats pratiques	46
5	A propos du choix du critère et de l’algorithme d’optimisation	49
6	Conclusion	52

1 Introduction

1.1 La segmentation Bayésienne avec une loi a priori Markovienne

Nous nous plaçons dans un contexte Bayésien. Considérons une distribution de probabilité quelconque $P(\omega)$ sur un espace des configurations noté Ω . Nous définissons alors une fonction de coût entre deux configurations $R(\omega, \omega')$. L'estimation Bayésienne consiste alors à minimiser le risque Bayésien, c'est à dire à rechercher la configuration qui minimise l'espérance de la fonction de coût :

$$\hat{\omega} = \arg \min_{\omega'} \int_{\omega} R(\omega, \omega') P(\omega) d\omega. \quad (1)$$

Traiter un problème dans le cadre de l'estimation Bayésienne revient donc à définir :

- Une distribution $P(\omega)$ sur l'espace des configurations Ω
- Une fonction de coût $R(\omega, \omega')$
- Un algorithme de minimisation

Considérons maintenant le contexte de la segmentation d'images. Le champ des données est noté $X = (x_s)_{s \in S}$ où S représente la trame de l'image (ensemble des sites) et les différents x_s sont à valeur dans l'espace d'état L (ensemble des niveaux de gris). L'image segmentée est notée $Y = (y_s)_{s \in S}$, les différents y_s étant à valeur dans Λ (ensemble des classes). La modélisation Markovienne permet de définir une loi *a priori* sur l'espace des configurations Λ^S , notée $P(Y)$. D'autre part, une modélisation des données (texture, bruit,...) permet de définir une loi d'attache aux données sur l'espace des configurations L^S , notée $P(X|Y)$. L'inconnue du problème étant Y , nous recherchons à "optimiser" la loi *a posteriori* définie par :

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}. \quad (2)$$

X étant donné, nous optimisons en réalité :

$$P(X|Y)P(Y). \quad (3)$$

Le choix de l’algorithme d’optimisation et celui de la fonction de coût sont bien évidemment liés. Nous rappelons deux critères couramment utilisés que nous comparerons dans la suite de ce travail.

Le MAP (Maximum *A Posteriori*) correspond à la fonction de coût définie par :

$$R_{MAP}(\omega, \omega_I) = 1 - \delta_{\omega_I}(\omega), \quad (4)$$

où δ représente le symbole de Kronecker. L’optimum est alors atteint pour la (une) configuration qui maximise la loi. Un tel critère est difficile à atteindre et nécessite un algorithme coûteux en temps de calcul comme le recuit simulé par exemple. Des approximations de ce critère peuvent néanmoins être obtenues par des algorithmes déterministes.

Le MPM (Modes Marginales *a Posteriori*) correspond à la fonction de coût définie par :

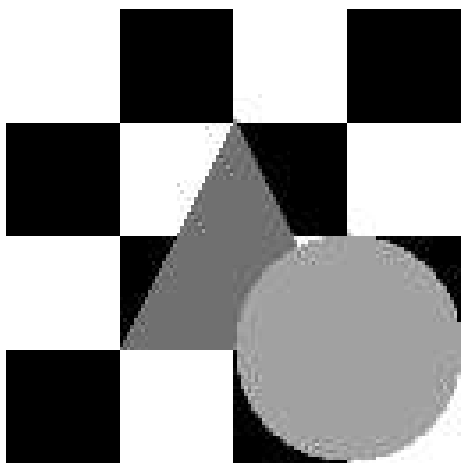
$$R_{MPM}(\omega, \omega_I) = \sum_{s \in S} 1 - \delta_{\omega_s I}(\omega_s). \quad (5)$$

L’optimum est atteint lorsque les différentes lois marginales en un site sont maximisées. Du point de vue algorithmique, les méthodes de type Metropolis-Hasting (Markov Chain Monte Carlo) sont utilisées pour optimiser un tel critère. Le critère MPM minimise alors le nombre de pixels mal classés et est donc adapté aux tâches de segmentation. En revanche, lorsque pour des contraintes de temps de calcul, une approximation du critère est recherchée par des algorithmes déterministes, le critère MAP est souvent préféré.

1.2 Présentation des images

Pour valider les méthodes et arguments proposés dans ce travail, nous considérons trois images tests. Les deux premières images sont des images de synthèse pour lesquelles nous connaissons donc “la solution” (voir figure 1). Nous pouvons alors déterminer le nombre de pixels mal classés et les localiser. La première contient des formes aux contours lisses alors que la seconde image contient une forme aux contours très chahutés, ce qui va permettre de tester la finesse des différents modèles de régularisation. Ces images sont bruitées par un bruit additif gaussien (voir figure 2). Pour finir, une troisième image permet de valider ce travail sur des données réelles. Il s’agit d’une image SPOT (canal

XS3) de la Hollande (voir figure 3). La scène représente une zone de parcellaire, possédant des structures très fines (routes, chemins, champs étroits,...).

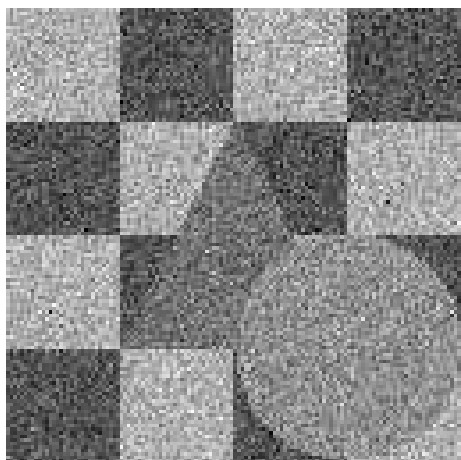


a) Image 256×256



b) Image 128×128

FIG. 1 – Images tests de synthèse



a) Image 1.a bruitée: SNB = 5dB



a) Image 1.b bruitée: SNB = 5dB

FIG. 2 – Images tests bruitées

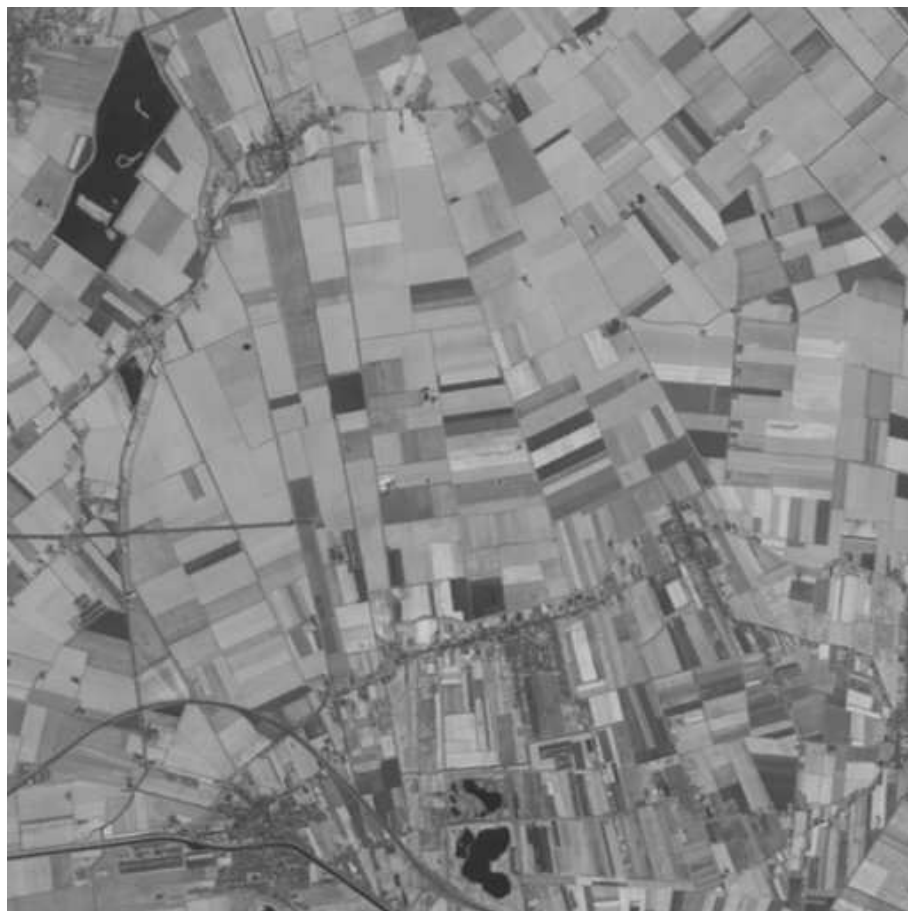


FIG. 3 – Image SPOT (canal XS3)

2 Un modèle a priori adapté aux images

Dans ce paragraphe, nous considérons le modèle Markovien le plus souvent choisi comme loi *a priori* pour un problème de segmentation : le modèle de Potts. Nous montrons les limites de ce modèle et son manque d'adéquation avec le résultat recherché. Nous proposons alors un second modèle (le Chien-modèle) permettant de contrôler certaines propriétés de la solution.

2.1 Limites du modèle de Potts

Nous considérons un modèle de Potts sans champ externe. Les cliques d'un tel modèle sont alors constituées de paires de pixels voisins. La loi d'un tel modèle s'écrit alors :

$$P(Y) = \frac{1}{Z(\beta)} \exp \left(-\beta \sum_{c=\{y_s, y_{s'}\}, c \in \mathcal{C}} \delta_{y_s=y_{s'}} \right), \quad (6)$$

où \mathcal{C} désigne l'ensemble des cliques.

Le modèle de Potts dépend donc d'un paramètre : β . La probabilité d'une configuration est d'autant plus élevée que le nombre de cliques homogènes (paires de voisins ayant le même label) est importante. En outre, la loi du modèle peut s'écrire sous la forme :

$$P(Y) = \frac{1}{Z(\beta)} \exp -\beta \#_Y, \quad (7)$$

où $\#_Y$ représente le nombre de cliques homogènes de la configuration Y .

Avec un tel modèle, nous pouvons donc espérer contrôler la longueur des contours ou encore la taille des régions des réalisations obtenues. En simulant différentes (un grand nombre !) réalisations du modèle, nous pouvons alors estimer la moyenne du nombre de cliques homogènes, notée $\langle \#_Y \rangle$, en fonction de la valeur du paramètre β . Cette courbe dépend de la taille de la réalisation et du nombre de classes considérées. Néanmoins, le comportement de la courbe reste identique. Un exemple de cette courbe est illustré sur la figure 4. La courbe présente une forte inclinaison autour de la valeur critique du paramètre [17], [33]. Sur la figure 5, des réalisations d'un modèle de Potts à 4

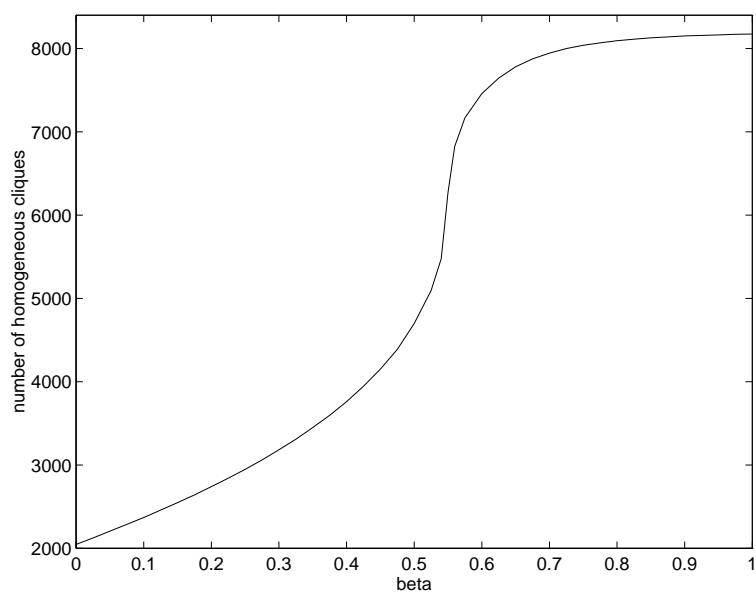


FIG. 4 – Nombre de cliques homogènes en fonction de β pour un modèle de Potts à 4 classes sur une trame 64×64

classes sont présentées pour différentes valeurs de β . Le modèle de Potts ne permet donc pas de modéliser des images comportant des zones homogènes du type des images segmentées désirées. En effet, pour des valeurs de β inférieures à la valeur critique, les réalisations obtenues sont fortement bruitées. Pour des valeurs supérieures à ce point critique, un des états devient rapidement prédominant et les réalisations sont constituées d'un fond homogène avec plus ou moins de bruit. Cette rupture de symétrie entre les différents états correspond à une transition de phase du modèle de Potts et a été abondamment étudiée en physique statistique [19], [15]. Du point de vue du traitement des images, elle signifie que le modèle de Potts est mal adapté pour générer des images formées de régions homogènes.

En pratique, ce modèle est malgré tout très utilisé comme terme régularisateur. Dans certains cas, les résultats sont acceptables car le terme d'attache aux données est suffisamment prépondérant et permet d'obtenir des régions homogènes. Ceci reste néanmoins valable uniquement si les données sont peu bruitées. Dans ce cas, l'utilisation d'une modélisation markovienne peut alors paraître inutilement sophistiquée. Lors de données bruitées, l'utilisateur est alors amené à utiliser une valeur de β nettement supérieure à la valeur critique, dès lors les structures fines de l'image sont perdues et la segmentation résultante est relativement grossière. Si nous considérons la segmentation recherchée comme la réalisation d'un modèle de Potts et estimons le paramètre associé par un critère de maximum de vraisemblance, la valeur estimée s'avère bien inférieure à la valeur utilisée en pratique pour obtenir une segmentation exempte de bruit. La figure 6 représente une photographie aérienne bruitée. Une segmentation manuelle est effectuée (voir figure 6.b). Cette segmentation est alors considérée comme la réalisation d'un modèle de Potts, ce qui conduit, par une estimation au sens du maximum de vraisemblance, à une valeur du paramètre β égale à 0.544. La figure 6.c représente une simulation du modèle de Potts avec le paramètre estimé. Nous pouvons remarquer que nous sommes loin d'une image formée de régions homogènes. La présence de bruit se retrouve alors sur le résultat de la segmentation obtenues avec une valeur de β effectivement égale à 0.544 (voir la figure 6.d). En prenant une valeur du paramètre β plus forte, nous obtenons alors une segmentation grossière où de nombreux détails ont été perdus (figure 6.e). En résumé, le modèle de Potts ne semble pas bien adapté à des tâches de segmentation d'images [33]. En

réalité, ce modèle ne fait pas la distinction entre contour et bruit. Avec ce modèle, régulariser signifie minimiser le nombre de cliques inhomogènes, c'est à dire minimiser la longueur des contours, ce qui n'est pas identique à minimiser le bruit. Un modèle possédant plusieurs paramètres est alors nécessaire pour contrôler de façons distinctes la longueur des contours et le bruit dans les réalisations. D'autre part, le comportement du nombre moyen de cliques homogènes en fonction du paramètre β autour du point critique est un inconvénient majeur du modèle. En effet, les images "interessantes", c'est à dire les images formées de régions homogènes correspondent, au sens de l'estimation de β , à des valeurs de β très proches les unes des autres (proches de la valeur du point critique). Le modèle de Potts permet donc en pratique de répartir les images en trois classes : le bruit, les images quasi-uniformes et "le reste du monde". La valeur du paramètre ne reflète donc aucune propriété intéressante de l'image. Pour affiner les résultats des segmentations d'image, il semble donc nécessaire de développer des modèles plus sophistiqués pour lesquels les différents paramètres permettent de classer les images suivant des propriétés déterminées comme la longueur des contours (interfaces entre plusieurs régions homogènes), la présence et la longueur des lignes (régions homogènes d'épaisseur 1 pixel) ou encore la taille moyenne des régions homogènes. Ces modèles nécessitent de considérer des cliques contenant plus de pixels pour lesquelles les différentes configurations sont réparties en plusieurs classes. Un premier modèle, fondé sur des cliques 3×3 est proposé dans [23] alors qu'un modèle sur une trame hexagonal est décrit dans [1]. Ces modèles considèrent les différentes configurations indépendamment les unes des autres. Nous proposons dans le paragraphe suivant de définir un modèle à partir des énergies locales, pour relier les différentes configurations formant un contour ou une ligne [30]. Une version 3D du modèle est présentée dans [11].

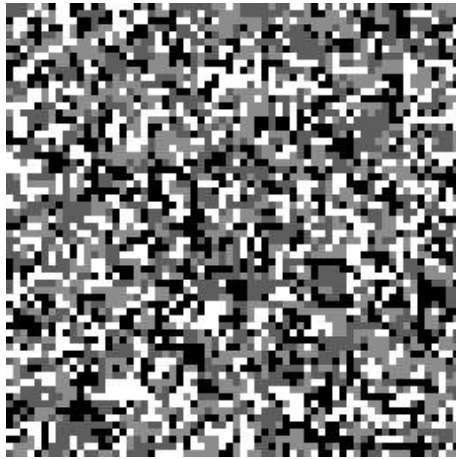
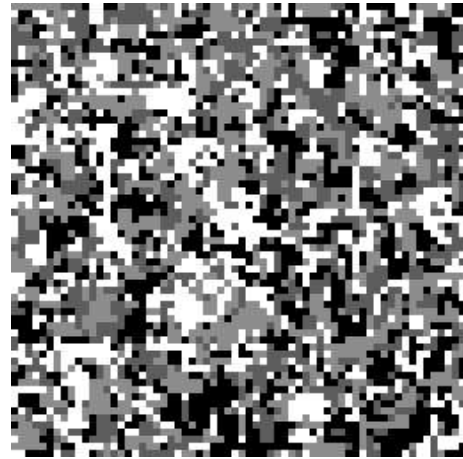
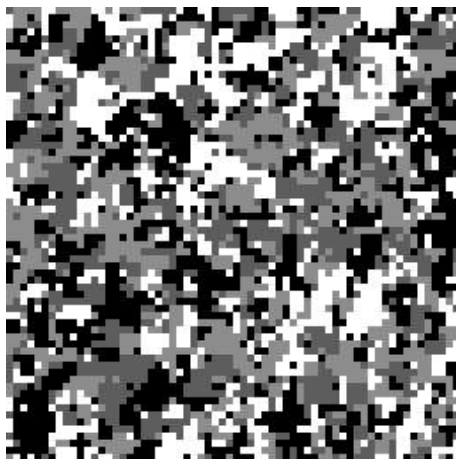
a) $\beta = 0.3$ b) $\beta = 0.4$ c) $\beta = 0.45$ d) $\beta = 0.55$

FIG. 5 – Réalisations du modèle de Potts

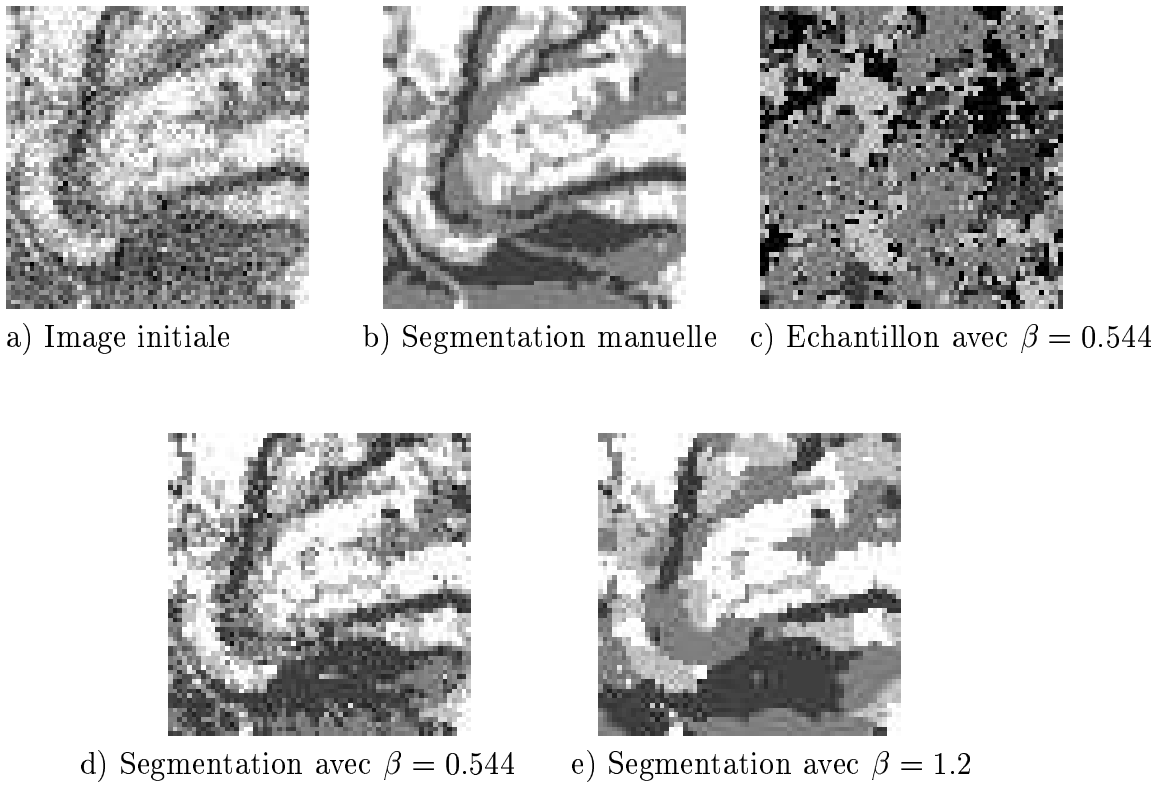


FIG. 6 – Segmentations avec le modèle de Potts

2.2 Le Chien-modèle

Le modèle présenté dans ce paragraphe dépend de trois paramètres respectivement notés e , l et n qui font référence aux contours (“edge”), aux lignes (“line”) et au bruit (“noise”). Le principe de construction du modèle se décompose en plusieurs étapes. Nous proposons en premier lieu un modèle binaire, étendu au cas m -aire par la suite. Nous choisissons un système de cliques : dans le cas présent, les cliques sont constituées des blocs carrés de trois pixels sur trois. Cette taille correspond à la taille minimale nécessaire à la définition de contours et de lignes. D’autre part, une taille plus grande accroît considérablement la complexité du modèle. Nous effectuons alors une classification des différentes configurations binaires possibles pour les cliques choisies. A chaque classe est alors associé un paramètre, valeur du potentiel correspondant. Le nombre des paramètres obtenus est encore trop grand pour être utilisable en pratique. En outre, le lien entre ces paramètres et les propriétés des réalisations n’est pas immédiat. La seconde étape consiste donc à relier ces paramètres avec des quantités facilement interprétables de l’image par l’intermédiaire d’une analyse des énergies locales.

2.2.1 Construction du modèle binaire

Considérons donc les $2^9 = 512$ configurations binaires possibles sur une clique de 3 pixels par 3. Une première classification est obtenue en considérant que deux configurations sont équivalentes si elles se déduisent l’une de l’autre soit par la symétrie noir-blanc, soit par une rotation. Nous obtenons alors 51 classes d’équivalence dont un représentant, et le nombre d’éléments de chaque classe, sont présentés sur la figure 7.

A chaque classe est associé un paramètre définissant la valeur du potentiel des configurations de la classe. Ainsi, sous les hypothèse d’isotropie du modèle et de symétrie noir-blanc, nous avons pour une telle topologie (cliques de 3×3 pixels) cinquante et un degrés de liberté. L’énergie étant définie à une constante près, nous imposons un potentiel nul aux configurations uniformes ($C(1) = 0$), ce qui ramène à cinquante le nombre de degrés de liberté. La construction du modèle consiste alors à imposer des contraintes en reliant les différents paramètres entre eux. Nous considérons que les états fondamentaux (états

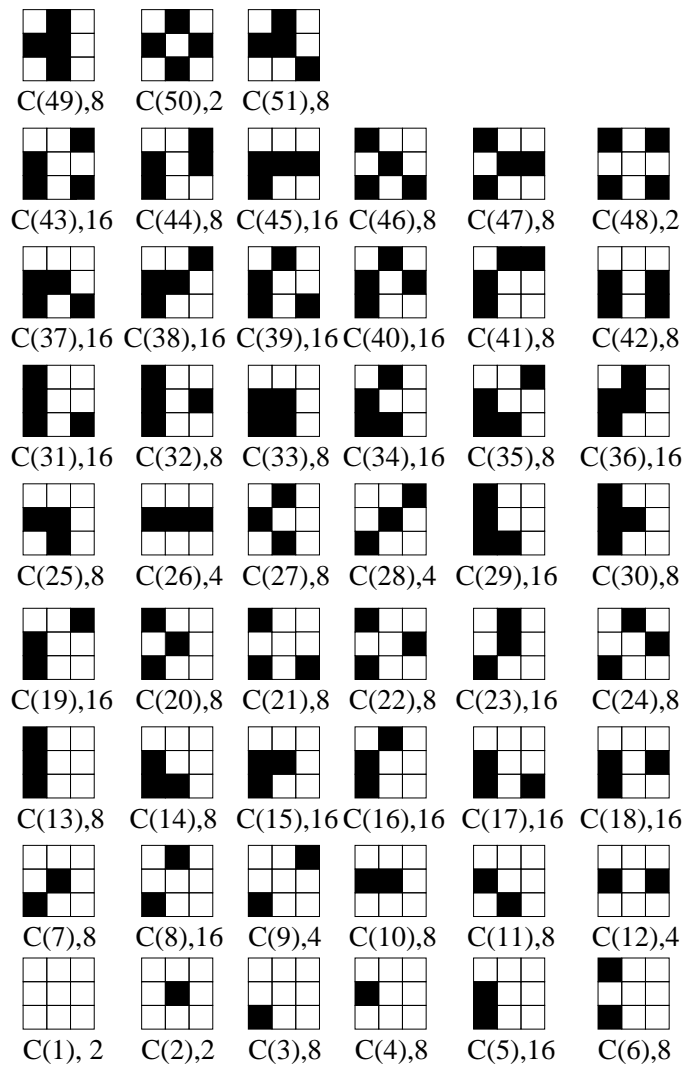


FIG. 7 – Classes d'équivalence des configurations binaires 3×3 et nombre d'éléments dans chaque classe

d'énergie minimale) sont données par (au moins) les configurations uniformes. De sorte que la méthodologie de construction du modèle peut se résumer ainsi :

- Modéliser les contraintes sur les réalisations globales par les énergies locales et en déduire les équations sur les paramètres définis par la figure 7.
- Vérifier la cohérence du système d'équations et le résoudre.
- Vérifier qu'aucune réalisation n'est d'énergie négative.

Le premier type de contraintes concerne la pénalisation des contours. Nous imposons le coût d'un contour par une énergie par unité de longueur, notée e . Avec la taille élémentaire choisie (taille des cliques), nous avons huit directions possibles (voir figure 8). En considérant le modèle symétrique par rotation, à ces directions correspondent trois types de contours, à savoir vertical, oblique et diagonal. Nous imposons à chaque direction une même énergie par unité de longueur pour obtenir un modèle isotrope. Au trois types de contours sont alors associées les trois équations suivantes, représentées sur la figure 9 :

$$2C(13) = e \quad (8)$$

$$2C(3) + 2C(14) = \sqrt{2}e \quad (9)$$

$$C(5) + C(29) + C(3) = \frac{\sqrt{5}}{2}e \quad (10)$$

Ces contraintes sont définies pour les objets de largeur au moins égal à 3. Pour les autres objets, nous définissons des contraintes similaires. Pour les objets d'épaisseur 1, nous introduisons un paramètre supplémentaire l déterminant l'énergie par unité de longueur des lignes, ce qui induit trois nouvelles équations (cf figure 10). Les objets d'épaisseur 2 et 3 sont définis comme ayant une énergie $2e$ par unité de longueur correspondant aux bords droit et gauche. Les équations obtenues s'écrivent alors :

- pour les objets verticaux (et horizontaux) :

$$2C(13) + C(26) = l \quad (11)$$

$$4C(13) = 2e \quad (12)$$

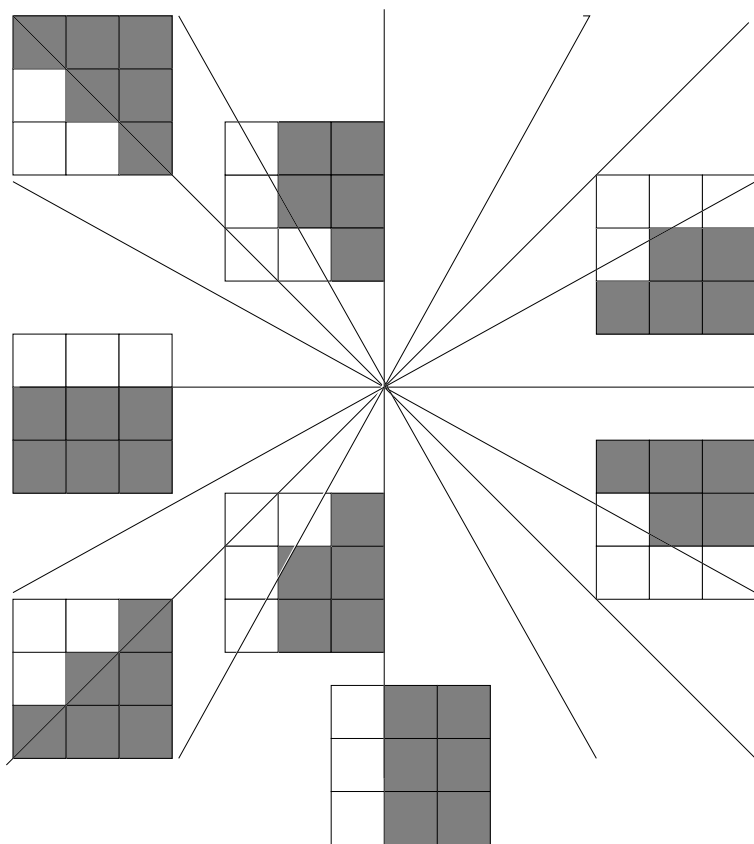


FIG. 8 – Orientations définies dans le chien-modèle

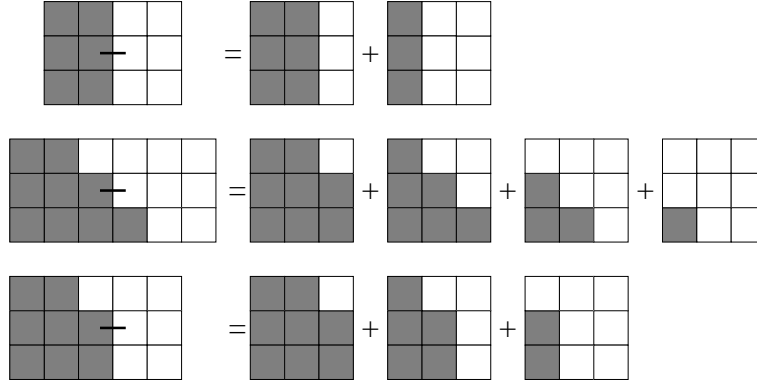


FIG. 9 – Equations correspondant aux contraintes de contours

– pour les objets diagonaux :

$$2C(3) + 2C(11) + C(28) = \sqrt{2}l \quad (13)$$

$$2C(3) + 2C(14) + 2C(35) = 2\sqrt{2}e \quad (14)$$

$$2C(3) + 4C(14) + C(9) = 2\sqrt{2}e \quad (15)$$

– pour les objets obliques :

$$C(3) + C(16) + C(23) + C(5) = \frac{\sqrt{5}}{2}l \quad (16)$$

$$C(3) + 2C(29) + C(19) + C(5) = \sqrt{5}e \quad (17)$$

$$2C(5) + 2C(29) + 2C(3) = \sqrt{5}e \quad (18)$$

Ces différentes contraintes se ramènent à onze équations faisant intervenir quatorze paramètres. Il nous faut maintenant vérifier que le système possède bien des solutions. Deux paramètres ($C(3)$ et $C(5)$) jouent un rôle particulier dans ce système puisqu'ils interviennent dans la plupart des équations. Nous exprimons alors les équations en fonction de ces deux paramètres pour obtenir le système suivant :

$$C(13) = \frac{e}{2} \quad (19)$$

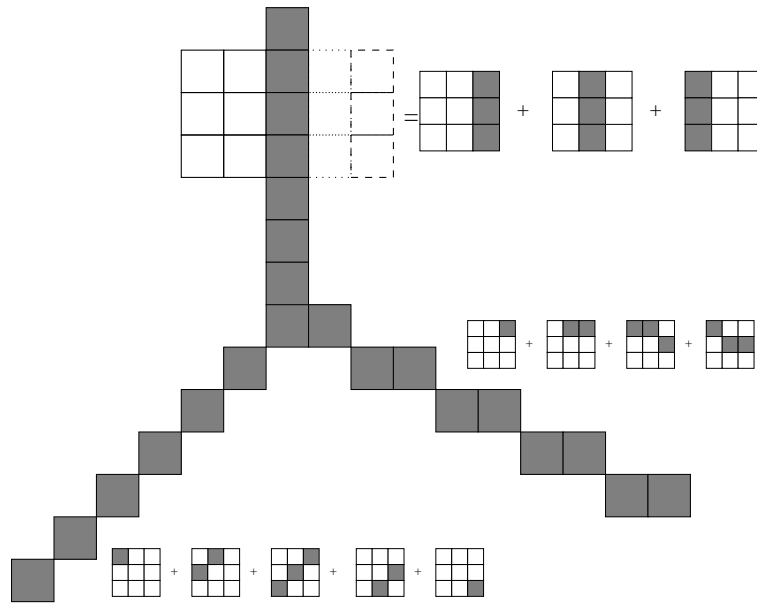


FIG. 10 – Définition des lignes dans le chien-modèle

$$C(14) = \frac{\sqrt{2}}{2}e - C(3) \quad (20)$$

$$C(26) = l - e \quad (21)$$

$$C(29) = \frac{\sqrt{5}}{2}e - C(3) - C(5) \quad (22)$$

$$2C(11) + C(28) = \sqrt{2}l - 2C(3) \quad (23)$$

$$C(35) = \frac{\sqrt{2}e}{2} \quad (24)$$

$$C(9) = 2C(3) \quad (25)$$

$$C(16) + C(23) = \frac{\sqrt{5}}{2}l - C(3) - C(5) \quad (26)$$

$$C(19) = C(3) + C(5) \quad (27)$$

Nous résolvons ce système en nous assurant qu'aucune configuration globale n'est d'énergie négative. Pour ce faire, nous considérons une solution pour laquelle aucune configuration locale n'est négative. Les configurations n'appa-

raissant pas dans le système d'équations sont considérées comme définissant le bruit et ont donc une énergie égale à n qui définit le troisième paramètre du modèle.

Pour des valeurs strictement positives des paramètres, nous avons uniquement les deux configurations uniformes comme états fondamentaux. En revanche si $e = l = 0$, toute configuration composée de bandes parallèles ou d'objets circulaires est un état fondamental.

L'inconvénient de ce modèle par rapport au modèle d'Ising réside essentiellement dans sa complexité. L'utilisation de cliques 3×3 entraîne un voisinage de 5×5 pixels. Le temps de calcul nécessaire à l'optimisation du modèle est donc accru par rapport à un simple modèle d'Ising 4 ou 8 connexe. En revanche, le modèle proposé est beaucoup plus souple puisqu'il fait la distinction entre bruit, lignes et contours. Il représente donc un *a priori* plus fin sur la solution de la segmentation, ce qui sera démontré sur les résultats expérimentaux.

2.2.2 Extension à un modèle m-aire

Dans ce paragraphe, nous étendons le chien-modèle binaire à un modèle m-aire. De même que pour le cas binaire, nous définissons un coût énergétique pour les contours, les lignes et le bruit respectivement. Afin de généraliser le cas binaire, deux approches peuvent être envisagées. La première consiste à pénaliser le désordre, la seconde à favoriser l'ordre. Explicitons ces deux approches sur l'exemple de la figure 11. Dans la première approche, cette configuration sera pénalisée du fait des pixels de la partie droite. En revanche, elle sera favorisée dans la seconde approche en vertu de l'uniformité de la partie gauche. Dans la première approche, certaines configurations de forte énergie (phases désordonnées) peuvent représenter de profonds minima locaux et poser des problèmes pour l'algorithme d'optimisation. En effet, certaines configurations seraient de forte énergie mais nécessiteraient le changement de labels de plusieurs pixels pour décroître en énergie. Pour éviter d'avoir des minima locaux trop profonds, nous adoptons donc ici la seconde approche.

Soient X une configuration sur une clique définie au paragraphe 2.2.1 et $\lambda_i, i \in \{1, \dots, n\}$ les différents labels présents dans la configuration. Considérons les différentes configurations binaires X_i obtenues en associant l'état 1 à toutes les sites ayant pour label λ_i et l'état 0 aux autres sites. Le potentiel $V_b(X_i)$

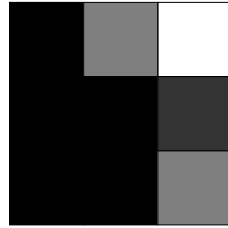


FIG. 11 – Configuration m-aire contenant une partie ordonnée et une partie désordonnée

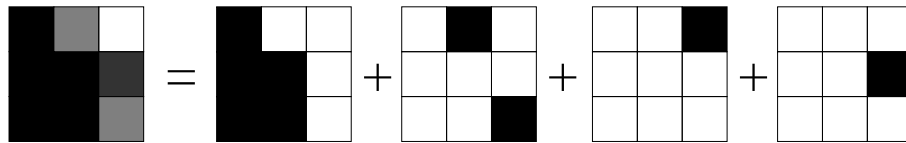


FIG. 12 – Généralisation du modèle binaire au cas m-aire

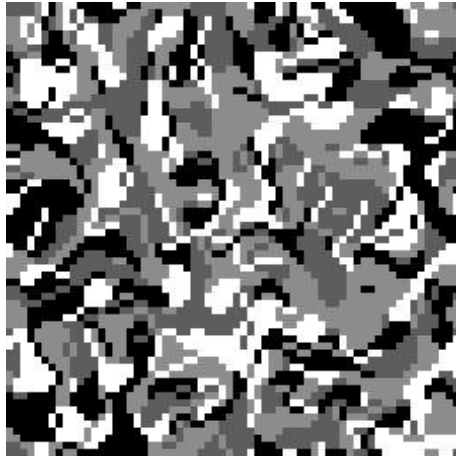
associé à la configuration X_i est alors celui correspondant au cas du modèle binaire. Le potentiel associé à la configuration X est alors donné par la formule (voir figure 12) :

$$V(X) = \sum_{i \in \{1, \dots, n\}} V_b(X_i). \quad (28)$$

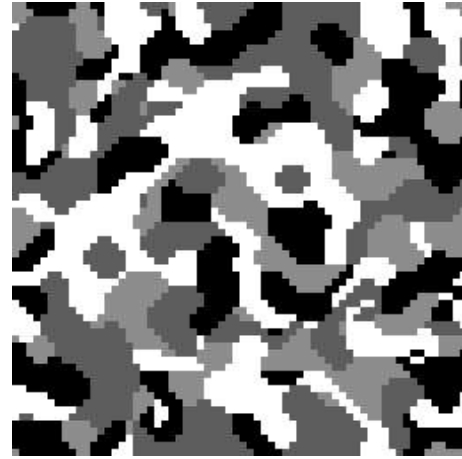
Sur la figure 13, nous présentons des simulations du chien-modèle pour différentes valeurs des paramètres. Bien que différentes des images réelles segmentées, ces images reflètent certaines propriétés structurelles (longueur des contours, nombre de lignes) permettant d'adapter les paramètres du modèle aux structures de l'image traitée, ce qui n'est pas le cas pour le modèle de Potts.

2.3 Résultat obtenus par les deux modèles

Dans ce paragraphe, nous comparons les deux modèles en utilisant une segmentation supervisée. Nous considérons comme connus le nombre de classes et les paramètres qui leur sont associés. En ce concerne l'image SPOT, des ensembles d'apprentissage ont été définis manuellement par un expert. Nous reprenons alors les paramètres donnés dans [18]. Pour chacun des deux modèles



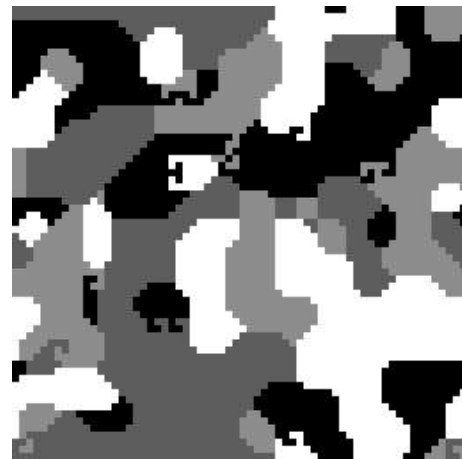
$e = 0.5, l = 1.2, n = 0.7$



$e = 0.7, l = 2.0, n = 0.7$



$e = 2.0, l = 4.0, n = 1.0$



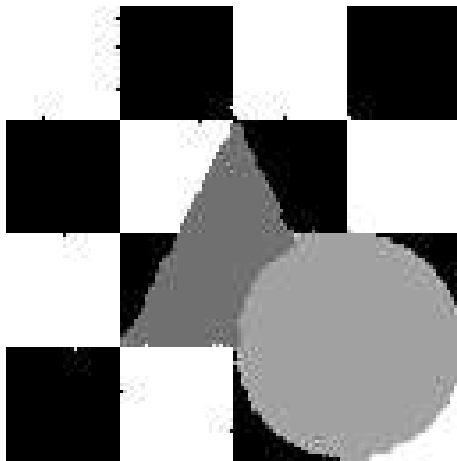
$e = 2.0, l = 6.0, n = 1.0$

FIG. 13 – Simulations du chien-modèle

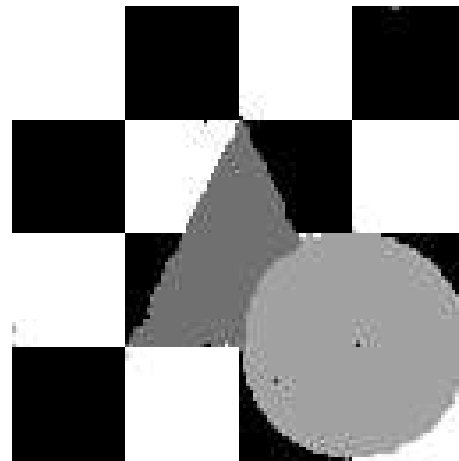
nous utilisons l'algorithme d'optimisation et le terme d'attache aux données qui donnent les meilleurs résultats et que nous expliciterons dans la suite de ce rapport.

La première image test donne des résultats comparables pour les deux modèles (voir figure 14). En effet, cette image ne possède pas de structures fines et peut donc être traitée correctement par le modèle de Potts. Il est intéressant de noter que le pouvoir régularisateur du chien-modèle est aussi efficace que celui du modèle de Potts. Pour finir, notons que le chien-modèle conduit à 85 pixels mal classés alors que nous en avons 99 pour le modèle de Potts. La différence provient essentiellement des contours de l'échiquier. La seconde image test présente une forme aux contours très accidentés. Le modèle de Potts a tendance à lisser les contours pour minimiser leur longueur, ce qui conduit à 342 pixels mal classés (voir figure 15.a). Le chien-modèle régularise sans trop pénaliser les contours et permet alors de réduire le nombre de pixels mal classés à 299 (voir figure 15.b).

Pour finir, nous présentons les résultats obtenus sur l'image SPOT. Les classes ont été définies à partir d'ensembles d'apprentissages déterminés par un expert. Ces ensembles concernent essentiellement le quart supérieur gauche de l'image. Les résultats sont donc surtout pertinents dans cette zone. Nous pouvons remarquer sur les figures 16 et 17 que le chien-modèle préserve mieux les structures fines comme par exemple les chemins entre les différents champs.



a) Modèle de Potts



b) Chien-modèle



c) Modèle de Potts : image des erreurs



d) Chien-modèle : image des erreurs

FIG. 14 – Résultats obtenus pour l'image de synthèse # 1



a) Modèle de Potts



b) Chien-modèle



c) Image des erreurs de a



d) Image des erreurs de b

FIG. 15 – Résultats obtenus pour l'image de synthèse # 2

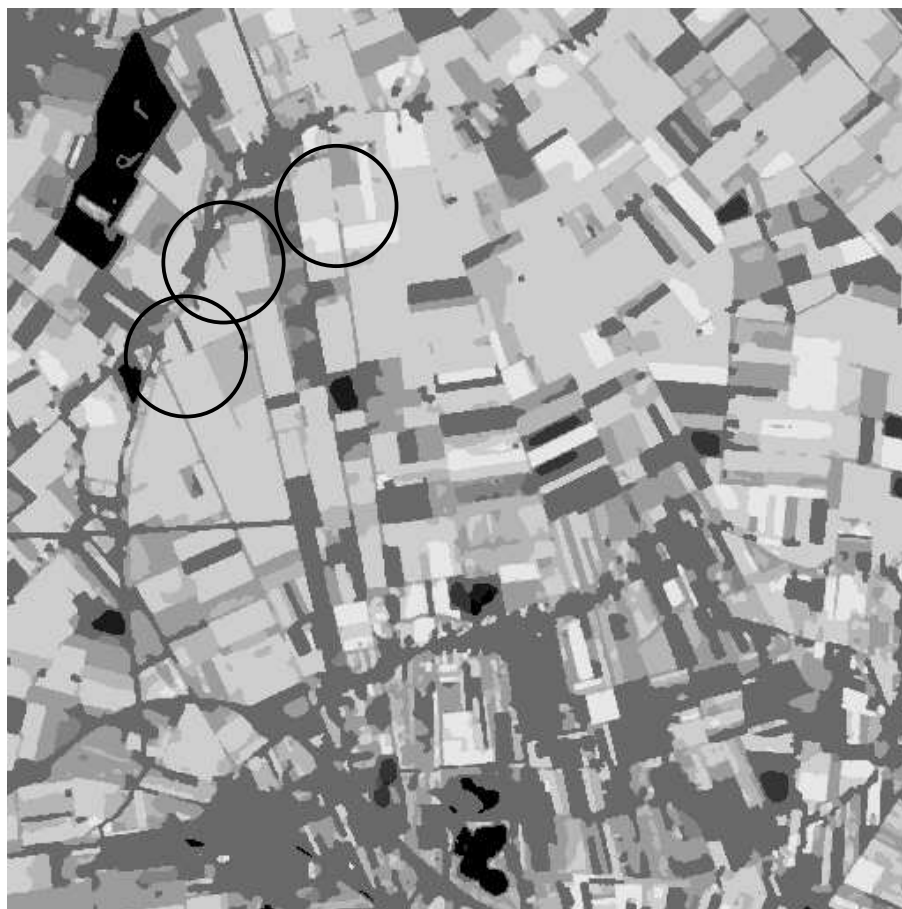


FIG. 16 – Segmentation supervisée de l'image SPOT par le modèle de Potts

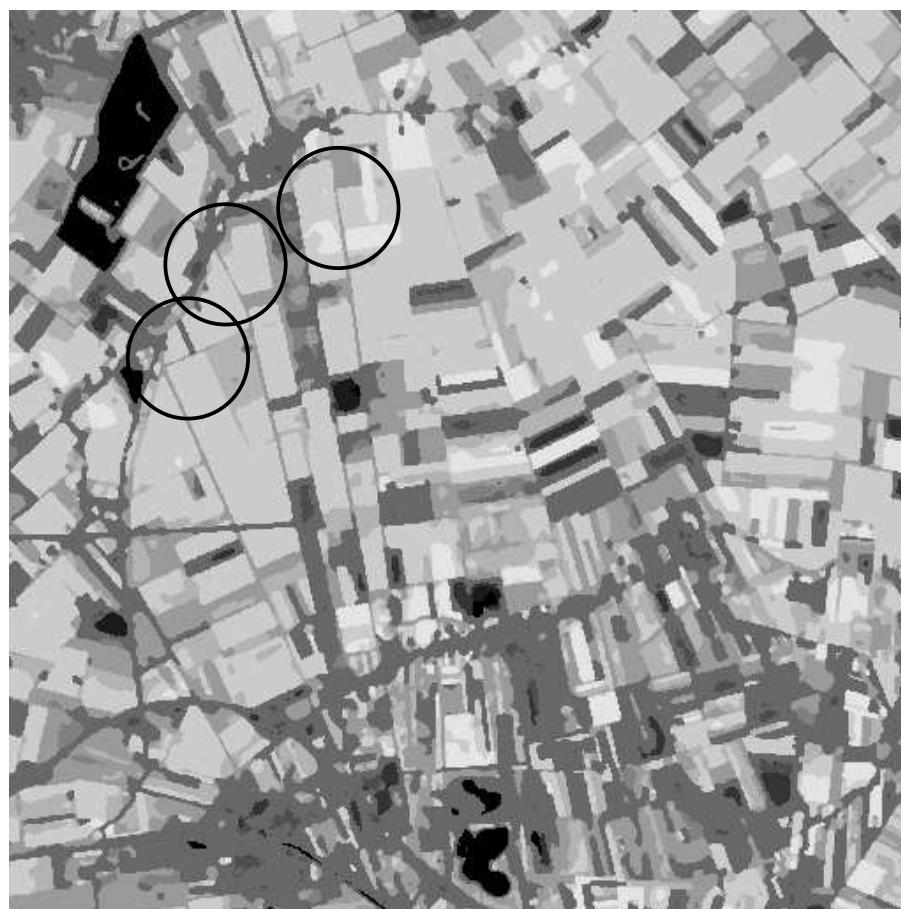


FIG. 17 – Segmentation supervisée de l’image SPOT par le chien-modèle

3 Estimation du modèle d'attache aux données

Dans ce paragraphe, nous considérons les problèmes du critère et de l'algorithme d'optimisation résolus. Pour effectuer la segmentation de façon non-supervisée, il reste alors à estimer les paramètres du modèle. Ces paramètres se divisent en deux types : ceux associés au terme d'attache aux données et ceux relevant du modèle *a priori*. Nous proposons dans ce paragraphe une méthode d'estimation du modèle d'attache aux données. Plusieurs méthodes d'estimation des autres paramètres du modèle, c'est à dire ceux de la loi *a priori* ont été proposées. Les plus répandues sont les méthodes de codages [2] pour des topologies particulières ou bien celles utilisant un critère de pseudo-maximum de vraisemblance dans un cadre plus général [3], [13], [9]. Des recherches en cours utilisant un critère de maximum de vraisemblance ("Markov Chain Monte Carlo Maximum Likelihood", [31]) feront l'objet d'un rapport ultérieur.

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. Nous proposons un algorithme pour estimer le nombre des classes et les paramètres qui leur sont associées. Le problème se ramène alors à l'estimation des composantes d'un mélange de distributions [20], [27].

3.1 Méthode de l'Histogramme de la Variance Conditionnelle (HVC)

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 [11].

3.1.1 Principe

Considérons un mélange de deux distributions P_1 and P_2 , dont les moyennes (resp. variances) sont notées μ_1 et μ_2 (resp. σ_1 et σ_2). Supposons que la probabilité de la classe 1 soit donnée par π 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 18.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 [14]. 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 ω_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 :

$$\begin{aligned}
 P(s \in C_1 | \omega_s = x) &= \frac{P(\omega_s = x | s \in C_1) P(s \in C_1)}{P(\omega_s = x | s \in C_1) P(s \in C_1) + P(\omega_s = x | s \in C_2) P(s \in C_2)}, \\
 P(s \in C_2 | \omega_s = x) &= \frac{P(\omega_s = x | s \in C_2) P(s \in C_2)}{P(\omega_s = x | s \in C_1) P(s \in C_1) + P(\omega_s = x | s \in C_2) P(s \in C_2)}, \quad (29)
 \end{aligned}$$

ce qui peut encore s'écrire :

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

$$A_2(x) = P(s \in C_2 | \omega_s = x) = \frac{(1 - \pi)P_2(x)}{\pi P_1(x) + (1 - \pi)P_2(x)}. \quad (30)$$

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 $\mu_{vois}(s)$, moyenne des voisins de s . La moyenne $\Lambda(x)$ de cette variable est donnée par :

$$\Lambda(x) = \langle \mu_{vois}(s) \rangle = A_1(x)\mu_1 + A_2(x)\mu_2. \quad (31)$$

Si tous les pixels de niveau de gris égal à x appartiennent à la classe 1 (resp. classe 2), 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. $\Lambda(x) = \mu_2$). Par conséquent, si il n'y a pas de conflit entre les deux classes, nous avons soit $A_1(x) = 0$ et $A_2(x) = 1$ ou bien $A_1(x) = 1$ et $A_2(x) = 0$. 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 18.b).

La variance $\Sigma^2(x)$ de $\mu_{vois}(s)$ est, quant à elle, donnée par :

$$\Sigma^2(x) = \frac{1}{N(x)} \sum_{s:\omega_s=x} \mu_{vois}(s)^2 - (A_1(x)\mu_1 + A_2(x)\mu_2)^2, \quad (32)$$

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

Alors :

$$\begin{aligned} \Sigma^2(x) = \frac{1}{N(x)} & \left(\sum_{s:\omega_s=x, s \in C_1} \mu_{vois}(s)^2 + \sum_{s:\omega_s=x, s \in C_2} \mu_{vois}(s)^2 \right) \\ & - (A_1(x)\mu_1 + A_2(x)\mu_2)^2. \end{aligned} \quad (33)$$

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

$$\Sigma^2(x) = A_1(x)(\sigma_1^2 + \mu_1^2) + A_2(x)(\sigma_2^2 + \mu_2^2) - (A_1(x)\mu_1 + A_2(x)\mu_2)^2, \quad (34)$$

c'est à dire :

$$\begin{aligned} \Sigma^2(x) = & A_1(x)\sigma_1^2 + A_2(x)\sigma_2^2 + A_1(x)(1 - A_1(x))\mu_1^2 \\ & + A_2(x)(1 - A_2(x))\mu_2^2 - 2A_1(x)A_2(x)\mu_1\mu_2. \end{aligned} \quad (35)$$

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

$$\Sigma^2(x) = A_1(x)\sigma_1^2 + A_2(x)\sigma_2^2 + A_1(x)A_2(x)(\mu_1 - \mu_2)^2. \quad (36)$$

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 18.c).

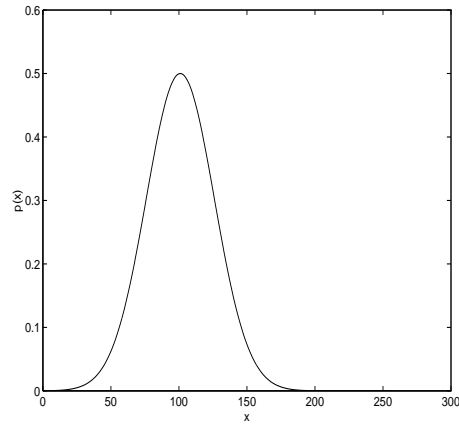
Entre deux classes successives, le long de la zone de conflit, nous avons :

$$\Sigma^2(\alpha) = \alpha\sigma_1^2 + (1 - \alpha)\sigma_2^2 + \alpha(1 - \alpha)(\mu_1 - \mu_2)^2, \alpha \in]0, 1[. \quad (37)$$

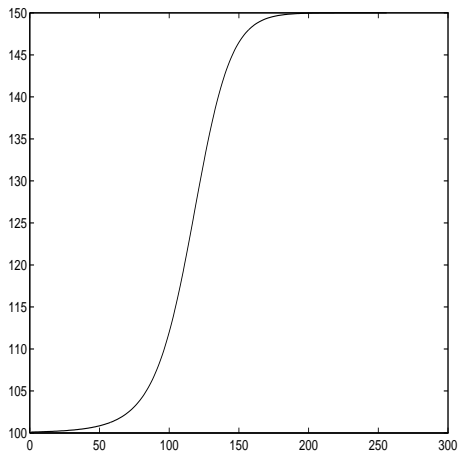
Cette fonction, dépendant de la variable α , est une parabole. Nous avons $\Sigma^2(0) = \sigma_2^2$ et $\Sigma^2(1) = \sigma_1^2$. Il est alors aisé de montrer que, si $|\sigma_1^2 - \sigma_2^2| < (\mu_1 - \mu_2)^2$, alors il existe $\alpha \in]0, 1[$ tel que, $\Sigma^2(\alpha) > \Sigma^2(0)$ et $\Sigma^2(\alpha) > \Sigma^2(1)$. Si ce maximum local est compris entre μ_1 et μ_2 , il permet de détecter la zone de conflit entre les deux classes (voir figure 18.c).

3.1.2 Détection des modes par une analyse d'un espace d'échelle

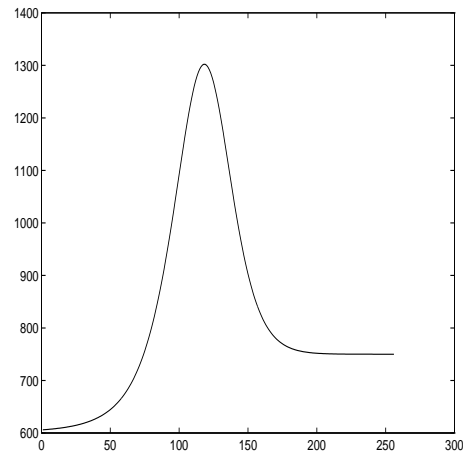
Nous présentons ici une méthode qui est notamment utilisée pour mettre en correspondance une courbe avec une somme de gaussiennes [22]. Un mode sur une courbe est défini comme un maximum local. Nous proposons d'utiliser une analyse d'un espace d'échelle, connue sous l'acronyme anglais SSA ("Scale Space 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 [32] a été appliquée pour détecter les modes de l'histogramme d'une image dans [4]. 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 [16] pour définir automatiquement le terme



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 $\Sigma^2(x)$

FIG. 18 – Détection d'un mélange par les moments conditionnels

d'attache aux données d'une modélisation markovienne, utilisée pour segmenter l'ALD (Adrénoleukodystrophy) à partir d'échos IRM (Images à Résonance Magnétique).

Pour illustrer notre propos, considérons une distribution gaussienne, caractérisée par sa moyenne μ et sa variance σ^2 (voir figure 19.a). Si nous convoluons cette distribution par un filtre gaussien centré, caractérisé par sa variance σ_f^2 , nous obtenons une gaussienne, dont la moyenne est donnée par μ 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 :

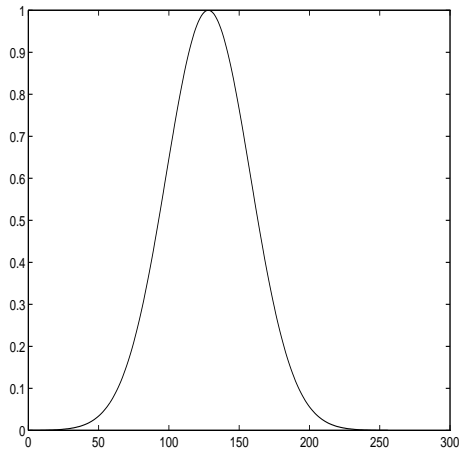
$$y = \mu \pm \sqrt{\sigma^2 + \sigma_f^2}. \quad (38)$$

Dans ce cas, l'espace d'échelle est donc constitué de deux courbes divergentes (voir figure 19.b).

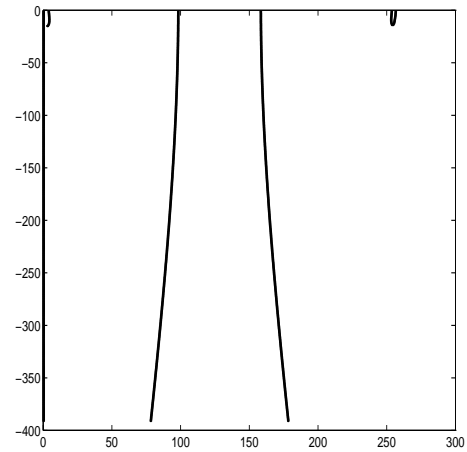
Considérons maintenant une distribution définie par la somme de deux gaussiennes (voir figure 19.c). La convolution par un noyau gaussien reste la somme de deux gaussiennes. Si les deux gaussiennes sont suffisamment séparées, nous obtenons une courbe bimodale et la dérivée seconde possède quatre zéros. 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. Lorsque σ_f^2 croît, les deux zéros centraux se rapprochent et finissent par se confondre pour disparaître (voir figure 19.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 [14].

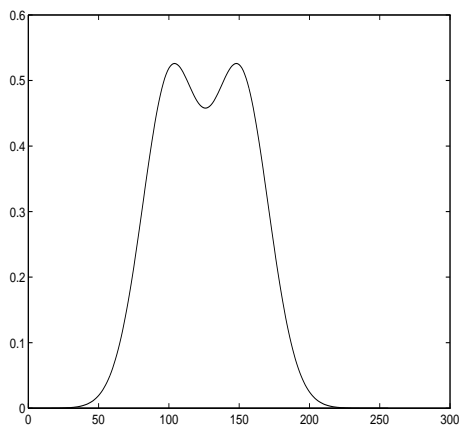
Nous proposons d'utiliser cette technique pour détecter les différents modes d'un HVC. Nous n'avons pas besoin d'une hypothèse gaussienne mais supposons simplement que chaque mode est caractérisé par un point d'inflexion de chaque côté. L'utilisation d'une SSA permet d'éviter de mauvaises détections dues au bruit. Considérons l'espace d'échelle associé avec un HVC donné. A



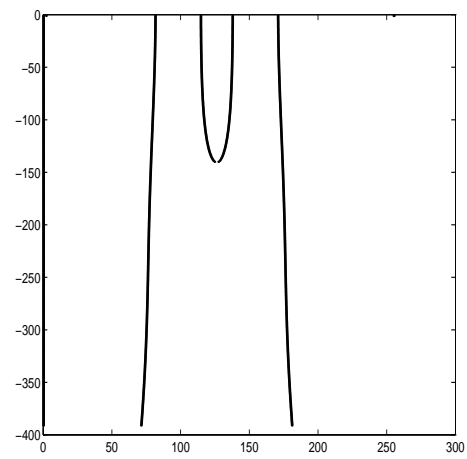
a) Gaussienne



b) Espace échelle associé à a



c) Somme de deux gaussiennes



d) Espace échelle associé à c

FIG. 19 – Gaussiennes et espaces échelle associés

partir de la convolution du HVC par un noyau gaussien de variance σ_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 20.

Considérons un cas particulier pour expliciter la méthode. Soit un HVC correspondant à quatre classes (voir figure 20.a). L'espace d'échelle associé et la définition des classes et des zones de conflit sont montrés sur la figure 20.b. Sur le sous ensemble noté B sur la figure 20.b, 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.

3.1.3 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 théorique 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

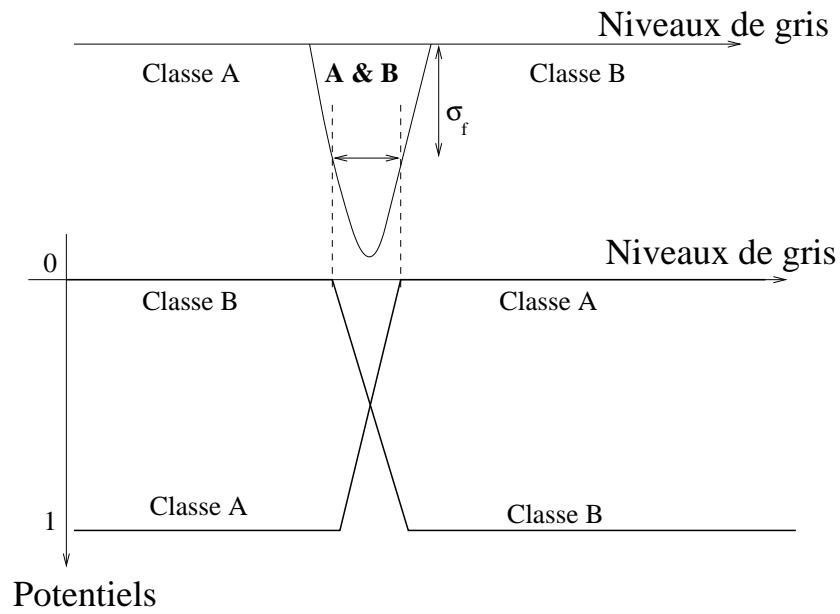
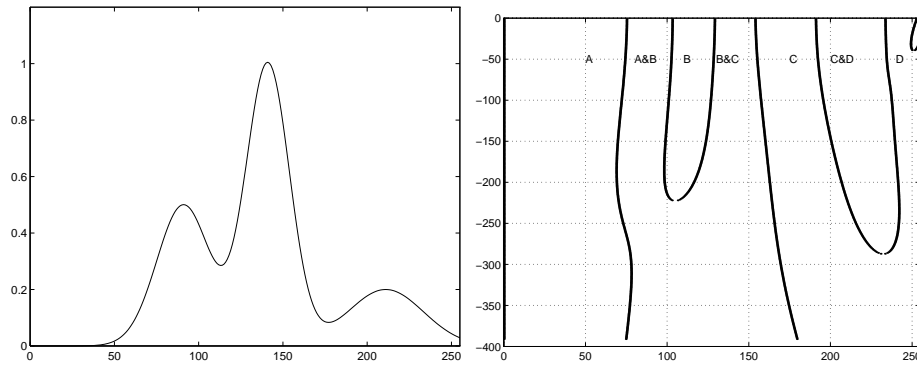


FIG. 20 – Définition des classes et des zones de conflit

ainsi démontrer l'apport de l'hypothèse contextuelle. Considérons un tel mélange défini par la probabilité *a priori* de la première classe π , les moyennes μ_1 et μ_2 , et les écarts type σ_1 and σ_2 . La loi du mélange s'écrit alors de la façon suivante :

$$p(x) = \pi p_1(x) + (1 - \pi)p_2(x), \quad (39)$$

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 :

$$\frac{d\Sigma^2(\alpha)}{d\alpha} = -2\alpha(\mu_1 - \mu_2)^2 + \sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2 = 0. \quad (40)$$

Le maximum est alors localisé par l'équation :

$$\text{Arg max } \Sigma^2(\alpha) = \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2}. \quad (41)$$

La fonction $\Sigma^2(\alpha)$ possède un mode entre μ_1 et μ_2 si et seulement si :

$$\begin{cases} \alpha(\mu_1) < \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \\ \alpha(\mu_2) > \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \end{cases} \quad \text{ou} \quad \begin{cases} \alpha(\mu_1) > \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \\ \alpha(\mu_2) < \frac{\sigma_1^2 - \sigma_2^2 + (\mu_1 - \mu_2)^2}{2(\mu_1 - \mu_2)^2} \end{cases} \quad (42)$$

En reprenant les notations de [20], nous définissons :

$$\Delta = \frac{\mu_2 - \mu_1}{\sigma_1} \quad \text{and} \quad \sigma = \frac{\sigma_2}{\sigma_1}. \quad (43)$$

Sans perte de généralité, nous prenons $\Delta \geq 0$ et $\sigma \leq 1$. La condition nécessaire et suffisante définie par (42) peut alors se résumer par les inégalités suivantes :

$$\begin{cases} \frac{1}{1 + (\frac{1}{\pi} - 1)\frac{1}{\sigma} \exp - \frac{\Delta^2}{2\sigma^2}} > \frac{1 - \sigma^2 + \Delta^2}{2\Delta^2} \\ \frac{1}{1 + (\frac{1}{\pi} - 1)\frac{1}{\sigma} \exp + \frac{\Delta^2}{2}} < \frac{1 - \sigma^2 + \Delta^2}{2\Delta^2} \end{cases} \quad (44)$$

En utilisant les notations suivantes :

$$E_\sigma = \frac{1}{\sigma} \exp - \frac{\Delta^2}{2\sigma^2} \quad \text{and} \quad E_1 = \frac{1}{\sigma} \exp + \frac{\Delta^2}{2}, \quad (45)$$

le domaine de validité est alors défini par :

$$\begin{cases} \frac{1}{\pi} < \frac{1}{E_\sigma} \frac{\Delta^2 - 1 + \sigma^2}{\Delta^2 + 1 - \sigma^2} + 1, \\ \frac{1}{\pi} > \frac{1}{E_1} \frac{\Delta^2 - 1 + \sigma^2}{\Delta^2 + 1 - \sigma^2} + 1. \end{cases} \quad (46)$$

Une condition nécessaire et suffisante de bimodalité dans le cas d'un mélange de deux gaussiennes peut être trouvée dans [20]. Cette condition s'exprime de la façon suivante. Considérons Δ_0 donné par :

$$\Delta_0 = \frac{1}{\sigma} \left[2(\sigma^4 - \sigma^2 + 1)^{3/2} - (2\sigma^6 - 3\sigma^4 - 3\sigma^2 + 2) \right]^{1/2}. \quad (47)$$

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

$$\pi_i^{-1} = 1 + \frac{\sigma^3 y_i}{\Delta - y_i} \exp \left(-\frac{1}{2} y_i^2 + \frac{1}{2} \frac{(y_i - \Delta)^2}{\sigma^2} \right), \quad i = 1, 2, \quad (48)$$

où (y_1, y_2) sont les racines de :

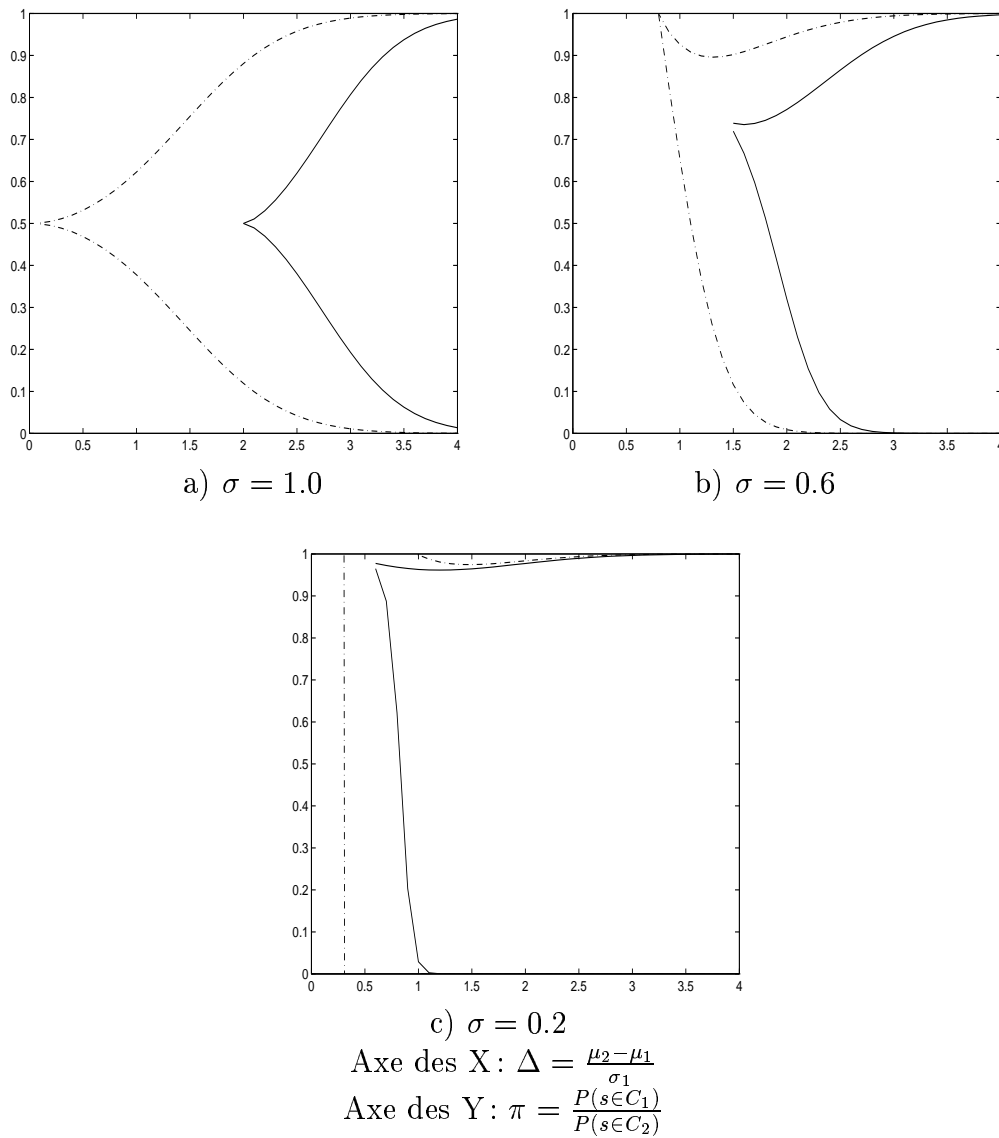
$$(\sigma^2 - 1)y^3 - \Delta(\sigma^2 - 2)y^2 - \Delta^2 y + \Delta\sigma^2 = 0, \quad (49)$$

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

La figure 21 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.

3.2 Principe des algorithmes de type EM et résultats

Pour étudier les performances de la méthode d'estimation proposée, nous l'insérons dans un algorithme de type EM ("expectation-maximisation") [29], [12]. Ce type d'algorithmes a été introduit en traitement des images à la fin des années 80 [5], [6], [7] et a fait l'objet de plusieurs variantes comme le

FIG. 21 – *Domaine de validité de la méthode HVC et domaine de bimodalité*

SEM [24] ou l'ICE [25], [26]. 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.

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 segmentation aboutissent aux mêmes conclusions que précédemment concernant les modèles *a priori*. Notons que l'apport du principe EM est le même pour les deux modèles. Néanmoins, le chien-modèle étant plus fin dans 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 23 et 24 montrent le résultat obtenus avec respectivement le modèle de Potts et le chien-modèle. Les structures fines sont nettement mieux préservées avec le chien-modèle.

Table 1

	HVC	P. (it. 1)	C. (it. 1)	P. (it. 2)	C. (it. 2)	P. (it. 3)	C. (it. 3)	V. R.
μ_1	70	84	78	84	82	84	84	85
σ_1^2	333	477	386	497	459	498	485	500
μ_2	107	113	107	114	111	114	113	115
σ_2^2	233	439	312	487	382	489	437	500
μ_3	147	145	145	144	144	144	144	145
σ_3^2	300	522	418	550	500	553	520	500
μ_4	185	175	177	174	175	174	175	175
σ_4^2	400	497	447	504	491	504	498	500

Estimations des paramètres associées 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.

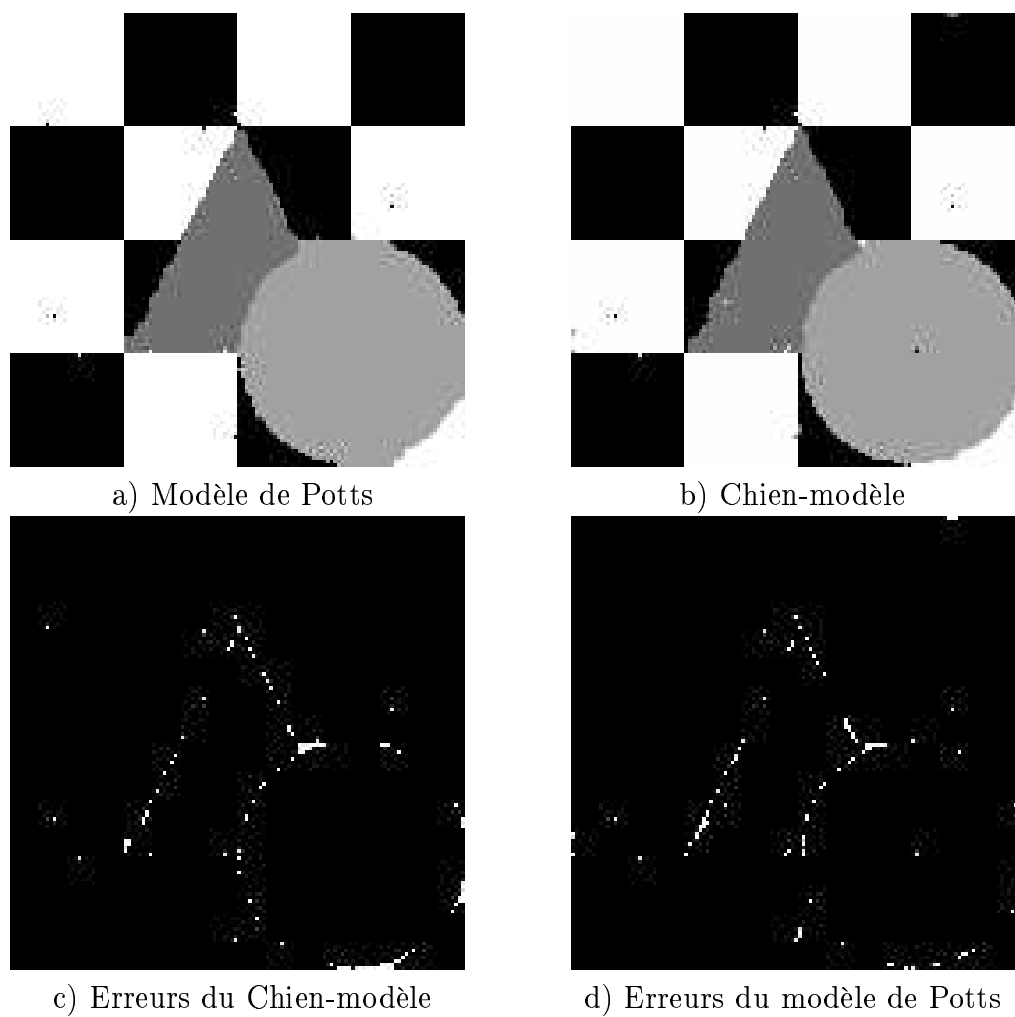


FIG. 22 – Résultats obtenus pour l'image de synthèse # 1 avec un ICE initialisé par un HVC

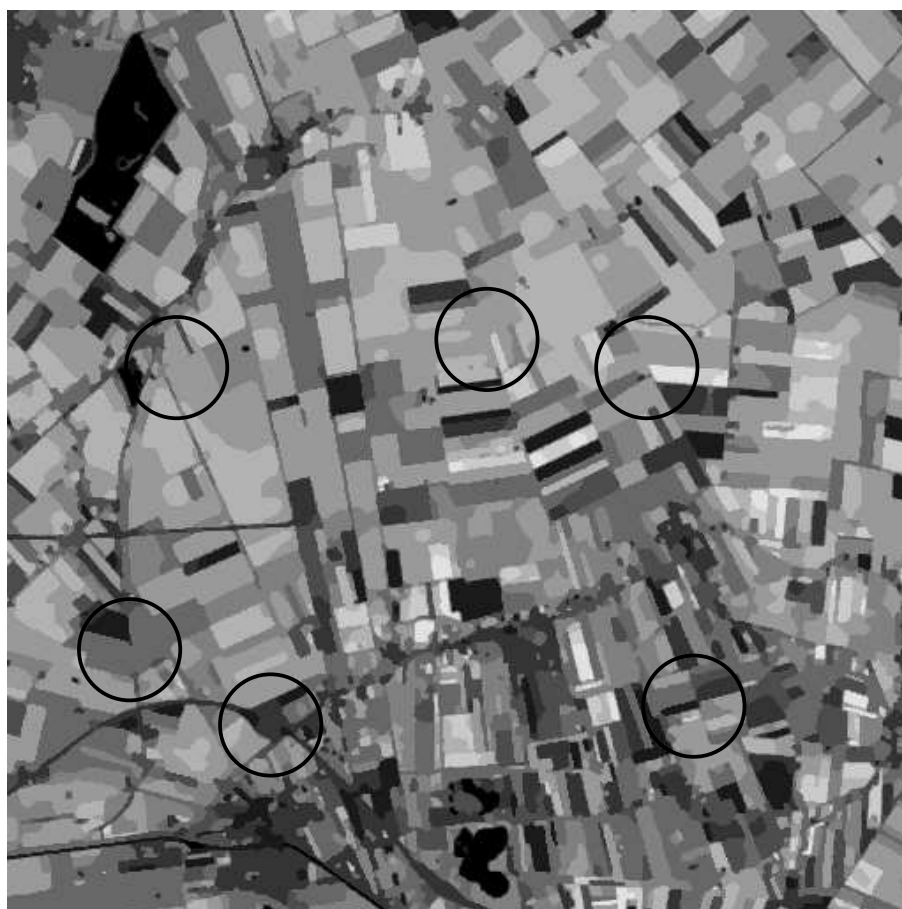


FIG. 23 – Résultats obtenus pour l'image SPOT avec un ICE initialisé par un HVC et un modèle de Potts



FIG. 24 – Résultats obtenus pour l'image SPOT avec un ICE initialisé par un HVC et un chien-modèle

4 A 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.

4.1 Entre loi bayésienne et relaxation de labels

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 [8], [10]. 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 :

$$U_{att}(X) = \sum_{s \in S} \sum_{\lambda \in \Lambda} (x_s - \mu_\lambda)^2 + \frac{1}{2} \log 2\pi \sigma_\lambda^2 \delta_{y_s = \lambda}. \quad (50)$$

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 labels 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 inhomogène [28]. 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 :

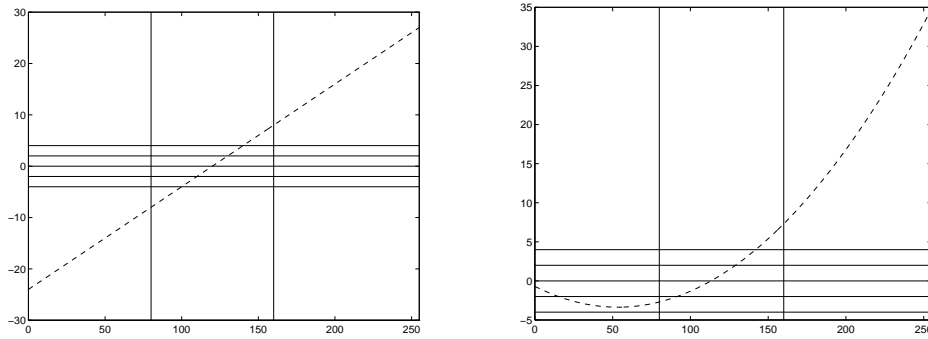
$$U_{att}(X) = \alpha \sum_{s \in S} \delta_{y_s = y_s^0}. \quad (51)$$

Pour étudier plus précisément le comportement de l'énergie associée à ces deux approches, nous considérons le cas d'un mélange de deux classes gaus-

siennes, 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 25 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 μ_1 et μ_2 correspond à ce que nous pouvons attendre. En effet, la différence entre les énergies d'attache aux données va croître entre μ_1 et μ_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 25). Par conséquent, si la valeur d'une donnée est extrême, 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 bornant de sorte qu'elle reste inférieure à l'énergie maximale du terme de régularisation.

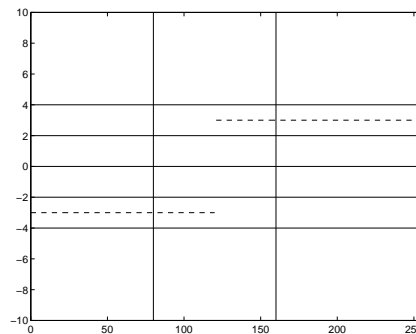
4.2 Influence sur les résultats pratiques

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 26 mettent partiellement en valeur ce phénomène. Néanmoins, pour ces images test le bruit considéré a été ajouté synthétiquement et suit donc réellement un loi gaussienne, ce qui limite la portée des remarques précédentes. En pratique, le modèle du bruit n'est qu'une approximation et certains pixels

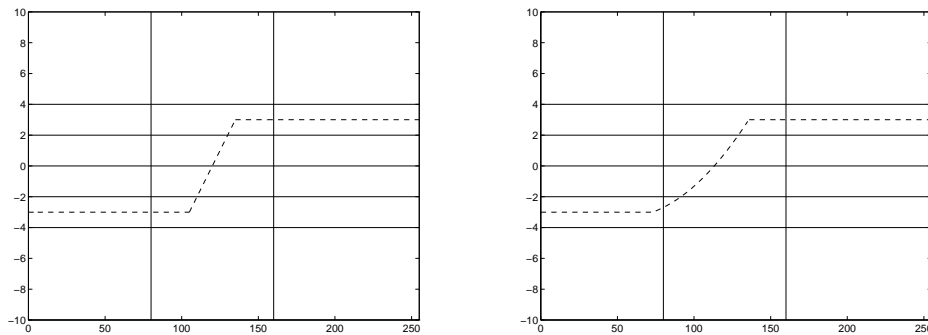


$\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 20$ $\mu_1 = 80, \sigma_1 = 20, \mu_2 = 160, \sigma_2 = 40$

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

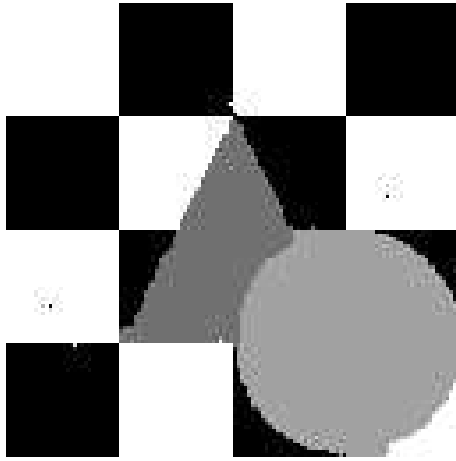
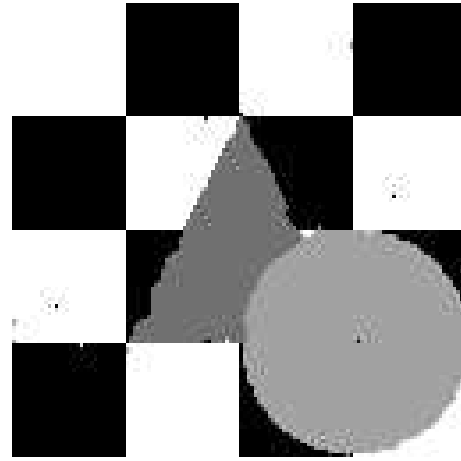
Courbes en pointillées : différence des énergies d'attache aux données $U_1 - U_2$

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

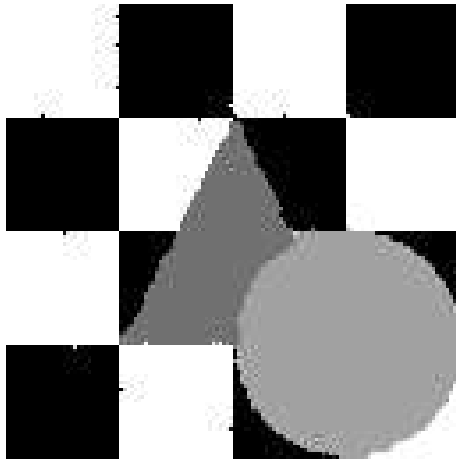
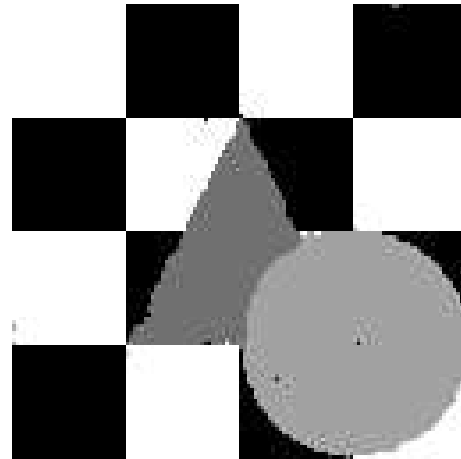
Les droites verticales représentent les moyennes μ_1 et μ_2

Axe des X : Niveaux de gris

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

c) Modèle de Potts: $\#E = 99$ d) Chien-modèle: $\#E = 85$

Attache aux données bornée

FIG. 26 – Influence du terme d'attache aux données

5 A propos du choix du critère et de l'algorithme d'optimisation

Les différents résultats présentés jusqu'ici ont été obtenus par un critère MPM. Pour converger vers ce critère, il suffit d'échantillonner le processus par un échantillonneur de Gibbs par exemple. Chaque pixel prend alors pour valeur le label le plus représenté en ce point dans les différents échantillons. Une alternative, couramment employée, consiste à estimer le MAP par un recuit simulé. Optimiser le MAP revient à rechercher une configuration d'énergie globale minimale. Néanmoins, les schémas théoriques de décroissance de la température sont beaucoup trop lents pour être envisageables en pratique. Une décroissance géométrique de la température est le plus souvent utilisée, ce qui conduit à une approximation du critère MAP. D'autres approximations sont également obtenues avec des algorithmes déterministes comme l'ICM ou le GNC. Le MPM semble plus intuitif puisqu'il permet de minimiser le nombre de pixels mal classés [21]. En revanche, le critère MAP peut être approché par de nombreux algorithmes déterministes peu coûteux en temps de calcul. Nous comparons ici quelques résultats obtenus avec les deux critères bayésiens : MAP et MPM.

Les résultats obtenus par les deux critères sont comparables lorsque le modèle de régularisation est le modèle de Potts (voir figure 27). En revanche, en utilisant le chien-modèle, nous avons une énergie plus complexe, comportant de forts minima locaux. Par suite, une décroissance trop rapide de la température lors d'un recuit simulé piège la configuration dans un minimum local. Le MPM est dans ce cas préférable à une approximation du MAP comme le montrent les résultats de la figure 28. Le chien-modèle montre donc encore sa supériorité mais est sensible en présence d'un fort taux de bruit aux approximations d'un algorithme.

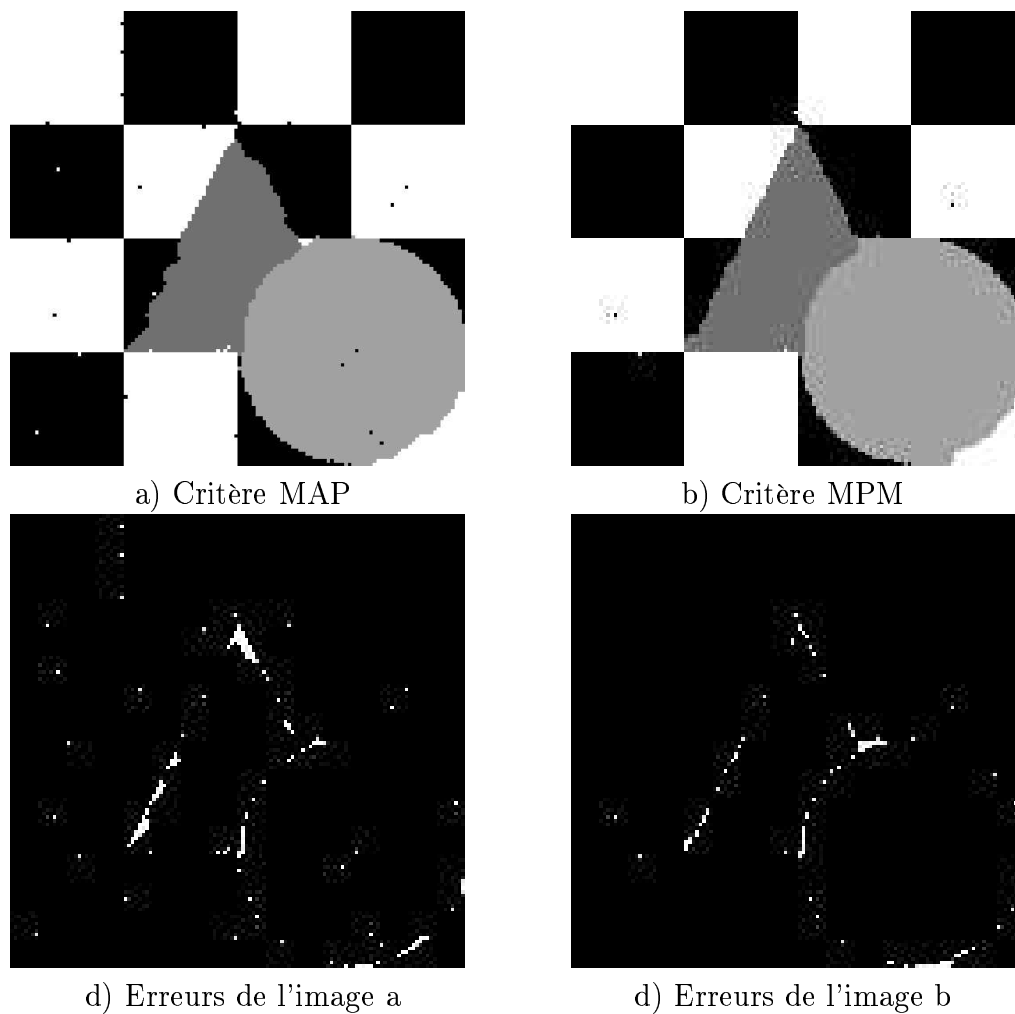


FIG. 27 – Résultats obtenus pour l'image de synthèse # 1 avec un modèle de Potts

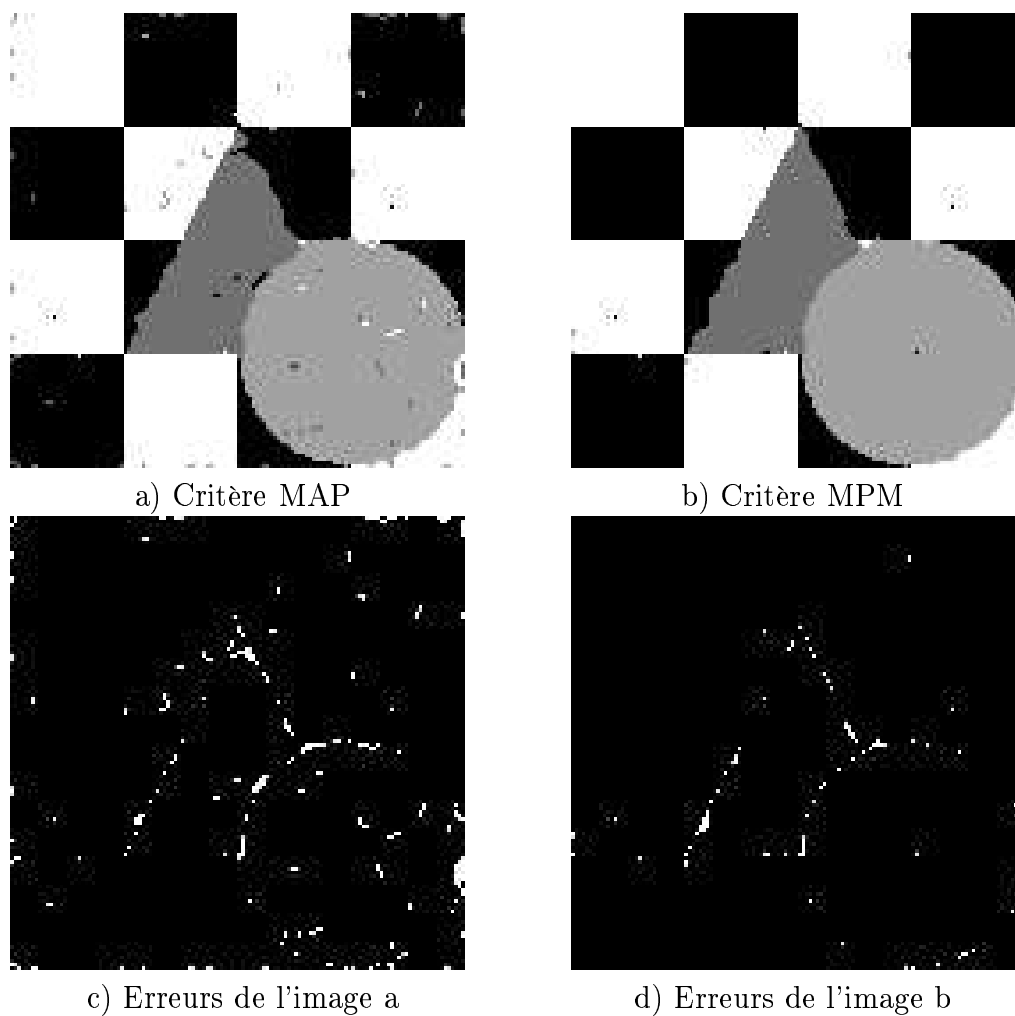


FIG. 28 – Résultats obtenus pour l'image de synthèse # 1 avec un chien-modèle

6 Conclusion

Dans ce rapport, nous avons revisité les différentes étapes de la segmentation d'images bayésienne fondée sur une modélisation markovienne. En premier lieu, nous avons défini un nouveau modèle *a priori* palliant les limites du modèle de Potts: le chien-modèle. Ce modèle, dont la version binaire est décrite dans [30], permet de contrôler l'énergie associée aux contours et aux lignes d'une réalisation indépendamment de l'énergie liée au bruit. Tout en conservant des propriétés régularisantes, ce modèle permet de préserver les structures fines de la segmentation. L'apport de ce modèle, comparativement au modèle de Potts, est montré à la fois sur des images de synthèse et sur des données réelles (imagerie SPOT).

Nous avons ensuite abordé le problème de l'estimation du nombre de classes et des paramètres associés à ces classes, permettant de définir le terme d'attache aux données. 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 [3], [13]. 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) [31]. Ce dernier point permettra d'obtenir une méthode totalement non-supervisée de segmentation.

Pour finir, nous avons comparé deux critères bayésiens (MPM et MAP) et analysé différents types d'attache aux données. Les résultats obtenus nous font préférer le MPM dont la convergence est mieux contrôlée du point de vue algorithmique que pour le MAP et une attache aux données intermédiaire entre celle donnée par la loi des classes et celle issue de la relaxation.

Références

- [1] H. Tjelmeland, J. Besag. Markov Random Fields with higher order interactions. submitted to JASA. Preprint.
- [2] J. Besag. Spatial interaction and statistical analysis of lattice systems. *Academic Royal Statistical Society Series B*, 36:721–741, 1974.
- [3] J. Besag. On the statistical analysis of dirty pictures. *J. Roy. Statist. Soc., Series B*, 48:259–302, 1986.
- [4] M.J. Carlotto. Histogram analysis using a Scale-Space approach. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 9(1):121–129, January 1987.
- [5] B. Chalmond. Image restoration using an estimated Markov model. *Signal Processing*, 15:115–129, 1988.
- [6] J. Zerubia, R. Chellappa. Mean field approximation using Compound Gauss Markov Random field for edge detection and image restoration. In *Proc. ICASSP, Albuquerque, USA*, 1990.
- [7] 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.
- [8] 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.
- [9] S. Lakshmanan, H. Derin. Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing. *IEEE trans. on Pattern Analysis and Machine Intelligence*, 11(8):799–813, August 1989.
- [10] X. Descombes. A fission and fusion Markovian approach for multi-channel segmentation. In *Proc. IGARSS'95 Firenze, Italy*, pages 124–127, 1995.
- [11] Xavier Descombes. Application of stochastic techniques in image processing for automatic tissue classification in MRI and blood vessel restoration in MRA. Technical Report KUL/ESAT/MI2/9603, Laboratory for Medical Imaging Research (ESAT-Radiology), K.U.Leuven, Belgium, 1996.

-
- [12] G. Celeux, J. Diebolt. The EM and the SEM algorithms for mixtures: statistical and numerical aspects. *Cahiers du CERO*, 32:135–151, 1991.
- [13] 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.
- [14] W. Press, S. Teukolski, W. Vetterling, B. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd ed., 1992.
- [15] H.O. Georgii. *Gibbs Measures and Phase Transitions*. De Gruyter - Studies in Mathematics, 1988. Vol. 9.
- [16] 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.
- [17] D. Higdon. *Spatial Applications of Markov Chain Monte Carlo for Bayesian Inference*. PhD thesis, University of Washington, 1994.
- [18] Zoltan Kato. *Modélisation markoviennes multir solutions en vision par ordinateur. Application la segmentation d'images SPOT*. PhD thesis, Université de Nice-Sophia Antipolis, December 1994.
- [19] S.K. Ma. *Modern Theory of Critical Phenomena*. The Benjamin/Cummings Publishing Company, Inc., 1976. Frontiers in Physics Lecture Note Series.
- [20] D.M. Titterton, A.F.M. Smith, U.E. Makov. *Statistical Analysis of finite mixture Distributions*. John Wiley and Sons, 1985.
- [21] J.L. Marroquin. *Probabilistic solution of inverse problems*. PhD thesis, MIT-Artificial Intelligence Lab., 1985.
- [22] A. Goshtasby, W.D. O'Neill. Curve fitting by a sum of Gaussians. *CVGIP: Graphical Models and Image Processing*, 56(4):281–288, 1994.
- [23] G. Wolberg, T. Pavlidis. Restoration of binary images using stochastic relaxation with annealing. *Pattern Recognition Letters*, 3:375–388, 1985.
- [24] 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.

-
- [25] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Traitement du signal*, 11(2), 1994.
- [26] Z. Kato, M. Berthod, J. Zerubia, W. Pieczynski. Unsupervised adaptative image segmentation. In *ICASSP'95, Detroit, Michigan, USA*, May 1995.
- [27] J. Diebolt, C.P. Robert. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B*, 56:363–375, 1994.
- [28] 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.
- [29] A.P. Dempster, N.M. Laird, D.B. Rubin. Maximum Likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. serie B*, 39(1):1–38, 1977.
- [30] X. Descombes, J.F. Mangin, E. Peckersky, M. Sigelle. Fine structures preserving model for image processing. In *Proc. 9th SCIA 95 Uppsala, Sweden*, pages 349–356, 1995.
- [31] C. J. Geyer, E.E. Thompson. Constrained monte carlo maximum likelihood for dependent data. *J. R. Statist. Soc. B*, 54(3):657–699, 1992.
- [32] A.P. Witkin. Scale space filtering: a new approach to multi-scale description. *Proc. Image Understanding*, pages 79–95, 1984.
- [33] R.D. Morris, X. Descombes, J. Zerubia. The Ising/Potts model is not well suited to segmentation tasks. In *IEEE Digital Signal Processing Workshop*, 1996. Sept. 1-4 1996 Norway.



Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENoble Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399