TREIZIÈME COLLOQUE GRETSI - JUAN-LES-PINS DU 16 AU 20 SEPTEMBRE 1991

SUR UNE NOUVELLE APPROCHE DE SEGMENTATION BAYESIENNE NON SUPERVISEE D'IMAGES

B. Braathen(1) N. Marhic(2) P. Masson(2) W. Pieczynski(3)

(1)Nordly Observatoriet
University of Tromso
N-9000 Tromso, NORVEGE

(2)Groupe Traitement d'Images
ENST Bretagne, BP 832
29285 Brest cedex, FRANCE

(3)Groupe Image, INT
9, rue C. Fourier
91011 Evry cedex, FRANCE

RÉSUMÉ

Cet article traite d'une approche non supervisée de la segmentation Bayesienne d'images. Le problème le plus important réside dans l'estimation d'un mélange de distributions. Pour le résoudre on utilise la procédure ICE, laquelle est décrite dans un cadre plus général. Trois algorithmes dérivés de cette procédure sont proposés dans le cas d'un mélange Gaussien. Leurs performances respectives sont comparées à celles d'algorithmes déjà existants.

ABSTRACT

This paper presents an unsupervised approach to the Bayesian segmentation of images. The main problem lies in the estimation of a distribution mixture. The ICE procedure provides a solution to this problem and after a recall about this procedure in a more general case, three algorithms are described in the case of a finite Gaussian mixture. Then, using simulations, their performances are compared with those of existing algorithms.

1 INTRODUCTION

Ce travail porte sur la segmentation non supervisée d'images. L'approche statistique adoptée utilise la modélisation stochastique suivante. S désignant un ensemble de pixels (ou sites), une image est modélisée par un couple $\{\zeta=(\zeta_s)_{s\in S}, X=(X_s)_{s\in S}\}$ de champs aléatoires. Chaque variable ζ_s est à valeurs dans un ensemble fini de classes $\Omega=\{\omega_1,..., \omega_k\}$ et chaque X_s est une variable aléatoire réelle. L'image que l'on observe, dite "réelle", est considérée comme une réalisation du champ X et le problème de la segmentation est celui de l'estimation de la réalisation invisible de ζ à partir des données X=x.

1.1 Segmentation Bayesienne

Il existe plusieurs approches Bayesiennes au problème de segmentation; en effet, de manière la plus générale, la segmentation Bayesienne consiste à estimer ζ_A par la configuration $\zeta_A{}^*$ dont la probabilité conditionnelle à $X_B=x_B$ (*probabilité a posteriori*) est maximale, ζ_A (resp. X_B) désignant la restriction de ζ (resp.X) à A (resp. B), avec A et B tels que $S\supseteq B\supseteq A$.

Parmi toutes les méthodes ainsi obtenues on peut distinguer deux groupes selon le choix fait pour B: celui des méthodes locales qui classent chaque pixel (A={s}) au vu des observations faites sur un voisinage V_s de ce pixel ($B=V_s$) et celui des méthodes globales qui utilisent toute l'information (B=S). Parmi ces dernières, on trouve le MPM [12] et le MAP [9] pour lesquelles A est pris, respectivement, égal à {s} et S. La mise en œuvre de l'une de ces méthodes, selon qu'elle appartienne à l'un ou l'autre groupe, requiert la connaissance de paramètres spécifiques parmi ceux définissant la loi de (ζ, X).

Dans la suite, nous désignerons par α et β, respectivement, les paramètres utiles relatifs à la distribution $P\zeta$ de ζ et ceux relatifs aux distributions conditionnelles $P_X{}^\varepsilon$ de X sachant $\zeta=\varepsilon$.

1.2 Segmentation non supervisée

Dans le cas de données réelles, la loi de (ζ, X) est le plus souvent inconnue et il se pose alors un problème d'estimation. La procédure ICE (Iterative Conditional Estimation) décrite dans [14] permet d'estimer de façon itérative le paramètre $\theta=(\alpha, \beta)$ nécessaire à la mise en œuvre de la segmentation. Le principe est le suivant. On définit un estimateur $\theta^*=\theta^*(\zeta, X)$ de θ à partir de (ζ, X). ζ étant inobservable, θ^* doit être approché par une fonction de X. La meilleure approximation, au sens des moindres carrés, est donnée par l'espérance conditionnelle à X. La procédure ICE se déroule alors de la façon suivante. Partant d'une valeur initiale $\theta^{(0)}$ et si $\theta^{(q)}$ est la valeur courante du paramètre, la nouvelle valeur est donnée par:

$$\theta^{(q+1)} = E_{(q)}[\theta^* | X] \tag{1}$$

où $E_{(q)}$ désigne l'espérance conditionnelle calculée en fonction de $\theta^{(q)}$.

Soit $P_{(q)}(\zeta | X)$ la distribution de ζ sachant X calculée sur la base de $\theta^{(q)}$, alors l'équation (1) peut s'écrire:

$$\theta^{(q+1)} = \sum_{\varepsilon} \theta^*(\varepsilon) P_{(q)}(\varepsilon|X) \tag{1'}$$

Le principe de cette procédure est différent du classique algorithme EM [5]; la notion de vraisemblance n'y intervient pas. Cependant, dans le cas simple d'un mélange Gaussien, les formules de réestimation du EM peuvent être obtenues à partir de (1) mais sous une fausse hypothèse [14].

1.3 Organisation de l'article

Dans la suite nous supposons X Gaussien conditionnellement à ζ. Nous décrivons, dans ce cas particulier, de nouveaux algorithmes dérivés de la procédure ICE et permettant d'estimer les paramètres utiles à la segmentation. Nous y rappelons aussi la démarche des algorithmes auxquels ils seront comparés. Le paragraphe suivant traite du cas local et le troisième du cas global. Les résultats sont présentés paragraphe 4 et le dernier paragraphe est consacré à la conclusion.

2 SEGMENTATION CONTEXTUELLE

La réalisation de chaque ζ_s est estimée à partir de $X_V=\{X_t : t \in V_s\}$ (la méthode correspondant à V={s} est dite "aveugle") en utilisant des fonctions discriminantes définies pour tout $i \in \{1,..., k\}$ par:

$$FD_i(x_V) = \sum_{j=1}^{k^{v-1}} P_{ij} f^{ij} \quad (2)$$

où v est le cardinal de V, $P_{ij}=P(\zeta_s=\omega_i, \zeta_{V^*}=\varepsilon_j^*)$ et f^{ij} est la densité de X_V sachant $(\zeta_s=\omega_i, \zeta_{V^*}=\varepsilon_j^*)$ avec $V^*=V-\{s\}$.

Alors:

$$\hat{\zeta}_s = \omega_i \Leftrightarrow FD_i(x_V) = \sup_{1\leq j\leq k} FD_j(x_V) \quad (3)$$

Supposons alors que $\zeta_V=\varepsilon_j$ avec la probabilité α_j et désignons par f_j la densité de X_V sachant $\zeta_V=\varepsilon_j$; f_j est Gaussienne de moyenne μ_j et de matrice de covariance Γ_j. Le calcul des fonctions discriminantes et, par suite, la segmentation sont possibles dès que l'on connaît $\alpha=\{\alpha_j\}_{1\leq j\leq m}$ et $\beta=\{\beta_j=(\mu_j,\Gamma_j)\}_{1\leq j\leq m}$ où $m=k^v$. Le plus souvent $\theta=(\alpha, \beta)$ est inconnu et on est alors confronté au problème de l'estimation des composants d'un mélange Gaussien.

2.1 Application de ICE

On considère une suite $V_1, V_2, ..., V_n$ de voisinages de forme V; en désignant par x_i la restriction de x à V_i on obtient alors un échantillon $(x_1, x_2, ..., x_n)$ de X_V.

La procédure ICE permet d'estimer θ à partir de cet échantillon. En effet, on peut définir un estimateur pour θ à partir de (ζ, X). α peut être estimé à partir des distributions empiriques et β par les moyennes et covariances empiriques.

Pour $i \in \{1,..., n\}$ et ε_j dans Ω^v, désignons par $p_{ij}^{(q)}$ la probabilité a posteriori, calculée sur la base de $\theta^{(q)}=(\alpha^{(q)}, \beta^{(q)})$, d'avoir la configuration ε_j sur V_i.

$$p_{ij}^{(q)} = p_{\theta^{(q)}}\{\zeta_i = \varepsilon_j | X_i = x_i\} \quad (4)$$

L'espérance conditionnelle $E_{(q)}[\alpha^* \mid X]$ peut être calculée et on obtient, pour chaque $\alpha_j^{(q)}$, la valeur moyenne des probabilités $p_{ij}^{(q)}$.

La loi des grands nombres permet d'approcher $E_{(q)}[\beta^* \mid X]$ par:

$$\frac{1}{N}\sum_{l=1}^{N} \beta^*(\xi_l, x) \quad (5)$$

où $\xi_1,..., \xi_N$ sont des réalisations de ζ selon les probabilités $p_{ij}^{(q)}$.

Nous désignerons donc par EMS l'algorithme ainsi obtenu; en effet, on peut remarquer que les formules de réestimation pour α sont celles de l'algorithme EM classique [15], la réestimation de β nécessitant des tirages stochastiques.

2.2 L'algorithme SEM

Cet algorithme est une variante de l'algorithme EM proposée par Celeux et Diebolt dans [3]. Il permet de définir une suite $(\theta^{(q)})_q$ en utilisant des tirages stochastiques. Les notations étant les mêmes que précédemment, le calcul de $\theta^{(q+1)}$ à partir de $\theta^{(q)}$ et $(x_1, x_2, ..., x_n)$ s'effectue en trois étapes:

(i) Calcul, pour chaque x_i, des probabilités a posteriori, $j \in \{1,...,m\}$, sur Ω^v.

(ii) Affectation de chacun des x_i à une "classe", qui est ici un élément de Ω^v, tirée dans Ω^v selon la loi de probabilité:

$$p_i^{(q)} = \{p_{ij}^{(q)}; j \in \{1,...,m\}\} \quad (6)$$

On obtient ainsi une partition $Q_1, Q_2,..., Q_m$ de l'échantillon $x_1, ..., x_n$.

(iii) On pose alors:

$$\alpha_j^{(q+1)} = \frac{1}{n}\mathrm{Card}(Q_j) \quad (7)$$

et chaque $\beta_j^{(q+1)}$ est obtenu en appliquant les estimateurs classiques, moyenne et covariance empiriques sur le sous-échantillon Q_j.

3 METHODES GLOBALES

On suppose que la loi de ζ est une distribution de Gibbs aux quatre plus proches voisins:

$$P_\zeta(\varepsilon) = K.\exp\{-U(\varepsilon)\} \quad (8)$$

où l'énergie U dépend d'un paramètre α, K étant la constante normalisatrice.

Les variables $(X_s)_{s\in S}$ sont supposées indépendantes conditionnellement à ζ. Dans ces conditions, la distribution a posteriori de ζ sachant X=x est aussi une distribution de Gibbs. On peut alors simuler des réalisations de ζ selon la distribution a posteriori pour autant que l'on connaisse $\theta=(\alpha, \beta)$, β étant le paramètre caractérisant le bruit. La segmentation par l'une ou l'autre des méthodes MAP et MPM est alors possible en utilisant, respectivement, l'algorithme du recuit simulé décrit par Geman *et al* dans [9] et celui proposé par Marroquin *et al* dans [12].

Le bruit étant supposé Gaussien on a $\beta=\{(\mu_i,\sigma_i)\}_{1\leq i\leq k}$, k étant le nombre de classes, et on s'intéresse à l'estimation de θ à partir de X=x.

3.1 Estimation des paramètres par ICE

Comme dans le cas contextuel, β est estimé par les moyennes et variances empiriques; le problème majeur réside dans le choix d'un estimateur pour le paramètre α caractérisant l'énergie de la distribution de Gibbs (8). La méthode de coding [2] a été très souvent utilisée mais elle requiert la résolution d'un système d'équations non linéaires. C'est pourquoi l'estimateur retenu est celui proposé par Derin *et al* dans [6]. L'énergie U peut s'écrire sous la forme d'une somme de fonctions, appelées potentiels, définies sur des ensembles, appelés cliques, formés d'un pixel seul ou d'un groupe de pixels mutuellement voisins. Les auteurs adoptent alors le modèle suivant (MLL). La valeur du potentiel correspondant aux cliques singletons est α_i si $\zeta_s=\omega_i$ et aux cliques d'ordre 2 est associé un paramètre α^j.

t' désignant l'ensemble formé des quatre plus proches voisins du pixel s, on peut écrire:

$$V(s,t',\alpha) = \phi^T(s,t')\ (\alpha_1,\ldots,\alpha_k,\alpha^1,\alpha^2)^T \qquad (9)$$

et, P(s, t') désignant la distribution jointe de s et t', on a, pour deux valeurs distinctes de s:

$$(\phi(i,t') - \phi(j,t'))^T\ \ \alpha = \mathrm{Ln}\left(\frac{P(j,t')}{P(i,t')}\right) \qquad (10)$$

et α^* est alors solution d'un système d'équations linéaires.

L'utilisation directe de (1) ou (1') est impossible. Cependant, en utilisant l'échantillonneur de Gibbs, on peut générer des réalisations $\xi_1,\ldots,\xi_N$ de ζ selon la distribution a posteriori et la loi des grands nombres permet les approximations suivantes:

$$\alpha^{(q+1)} = \frac{1}{N}\sum_{l=1}^{N}\alpha^*(\xi_l) \qquad (11.1)$$

$$\beta^{(q+1)} = \frac{1}{N}\sum_{l=1}^{N}\beta^*(\xi_l,x) \qquad (11.2)$$

Pour limiter le temps de calcul nous avons choisi N=1; on peut cependant remarquer que des valeurs supérieures pour N améliorent légèrement les résultats.

3.2 L'algorithme EM Gibbsien

Cet algorithme itératif, développé par B. Chalmond dans [4] combine l'algorithme MPM, pour la segmentation, avec l'algorithme EM, ce dernier permettant la réestimation des paramètres. Un voisinage de forme V_s a $m=k^{\mathrm{Card}(V_s)}$ configurations possibles. L'auteur suppose qu'à chacune de ces configurations correspond un "label"; alors le paramètre α caractérisant la distribution de ζ est donné par les probabilités conditionnelles:

$$p_{ij} = P(\varsigma_s = \omega_i | V_s \ \text{de type } j) \qquad (12)$$

Si f_θ désigne la pseudo-vraisemblance, définie à partir des densités conditionnelles des distributions a priori et a posteriori, alors l'algorithme EM définit $\theta^{(q+1)}$ par:

$$\underset{\theta}{\mathrm{Max}}\ E\left[\mathrm{Log} f_\theta(\varsigma,X) | X = x, \theta^{(q)}\right] \qquad (13)$$

On obtient alors les formules de réestimation suivantes:

$$p_{ij}^{(q+1)} = E\left[n_{ij}|x,\theta^{(q)}\right] \Big/ \sum_i E\left[n_{ij}|x,\theta^{(q)}\right] \qquad (14)$$

où n_{ij} est le nombre d'occurences dans ζ telles que $\zeta_s=\omega_i$ et V_s de type j.

$$\mu_i^{(q+1)} = \sum_{s\in S}\gamma_s^{(q)}(i)x_s \Big/ \sum_{s\in S}\gamma_s^{(q)}(i) \qquad (15)$$

et, σ étant supposé indépendant de la classe:

$$\sigma^{(q+1)} = \sum_i\sum_{s\in S}\gamma_s^{(q)}(i)(x_s - \mu_i^{(q+1)})^2 \Big/ \sum_i\sum_{s\in S}\gamma_s^{(q)}(i) \qquad (16)$$

où $\gamma_s^{(q)}(i)$ est la loi marginale de la distribution a posteriori:

$$\gamma_s^{(q)}(i) = P_{\theta^{(q)}}[\varsigma_s = \omega_i | X = x] \qquad (17)$$

On peut, à partir de simulations de la distribution a posteriori et à l'aide d'une méthode de Monte-Carlo approcher $E[n_{ij} \mid x, \theta^{(q)}]$ et (14). En maximisant cette dernière on obtient, à chaque itération, une segmentation par MPM.

3.3 L'algorithme de P.A. Devijver

Cet algorithme proposé par P.A. Devijver dans [7] et [8] pour la segmentation d'images utilise la modélisation par champs de Markov causaux.

4 RESULTATS

Nous avons appliqué les algorithmes décrits précédemment pour effectuer la segmentation de l'image 1. Celle-ci est une réalisation d'un champ Markovien, le modèle retenu pour le champ ζ étant celui de Ising.

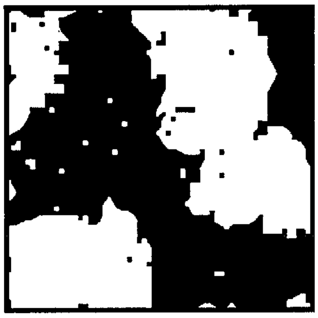


Image 1

En associant respectivement 0 et 2 aux classes, nous l'avons dégradée par addition d'un bruit blanc BB (resp. spatialement corrélé BC) Gaussien centré, de variance 1. De l'addition de ces bruits résultent respectivement l'image 2 et l' image 3.

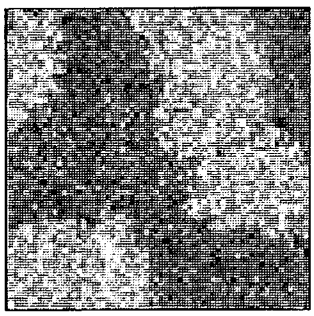


Image 2

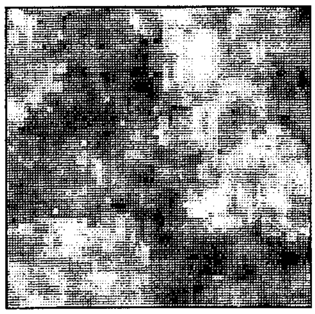


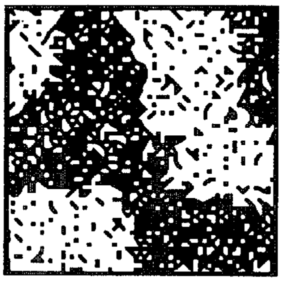
Image 3

La méthode d'initialisation retenue est celle de l'histogramme cumulé qui nous donne une première estimation des paramètres du bruit, laquelle permet une première segmentation par la méthode aveugle. On peut alors initialiser le paramètre α à partir de la segmentation obtenue. Les pourcentages de pixels mal classés sont donnés dans le tableau ci-dessous, MDV désignant la méthode "histogramme cumulé-classification aveugle".

bruit \ méthode	MDV	EMS	SEM	ICE MAP	ICE MPM	EM Gibbsien	Dev.
BB	16.1	6.4	4.3	3.6	2.4	2.2	2.6
BC	15.3	13.1	14.7	9.2	11.5	11.6	13.6

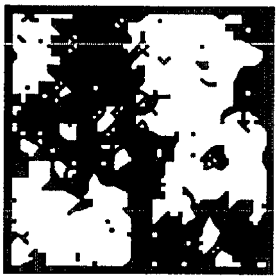
Tableau 1: pourcentages de pixels mal classés

Les segmentations obtenues sont présentées ci-dessous, la segmentation de l'image i par la méthode * étant désignée par *(image i) pour i=2 ou 3. Concernant les méthodes contextuelles (EMS et SEM) le voisinage V considéré est celui formé des quatre plus proches voisins.

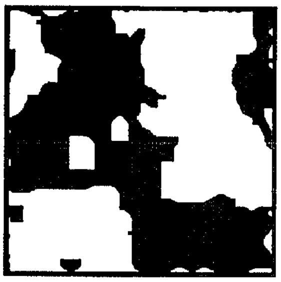

MDV (image 2)


MDV (image 3)

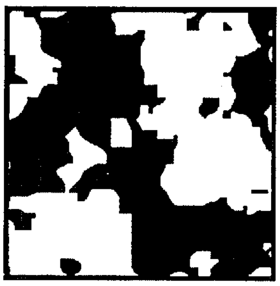

EMS (image 2)


SEM (image 3)


ICEMPM (image 2)


ICEMAP (image3)


DEV. (image 2)


EM Gibbsien (image 3)

5 CONCLUSION

La procédure ICE permet la conception de nombreux algorithmes, globaux ou locaux, de segmentation Bayesienne non supervisée. Dans le cas d'une image très homogène comme l'image 1, les performances des trois algorithmes proposés sont comparables à celles des algorithmes déjà existants; toutefois, la combinaison ICE-MAP s'avère la plus efficace dans la cas du bruit spatialement corrélé. Notre étude s'inscrit dans le cadre d'un projet plus général visant à établir un lien entre l'efficacité des différents algorithmes et certains facteurs, tels que l'homogénéité de l'image ou la corrélation du bruit.

REFERENCES

[1] R. Azencott, "Image analysis and Markov fields," *Proceedings of ICIAM*, Paris, 1987.

[2] J. Besag, "Spatial interaction and the statistical analysis of lattice systems," *Journal of the Royal Statistical Society, Series B*, 36, pp. 192-236, 1974.

[3] G. Celeux et J. Diebolt, "L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités," *Revue de Statistiques Appliquées, Vol. 34*, No.2, 1986.

[4] B. Chalmond, "An iterative Gibbsian technique for reconstruction of m-ary images," *Pattern Recognition, Vol. 22*, No.6, pp. 747-761, 1989.

[5] M.M. Dempster, N.M. Laird and D.B. Rubin, "Maximum likehood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B*, 39, pp. 1-38, 1977.

[6] H. Derin and H. Elliott, "Modelling and segmentation of noisy and textured images using Gibbs random fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-9*, pp 39-55, 1987.

[7] P.A. Devijver, "Image segmentation using causal Markov random field models," in *Pattern Recognition*, J. Kittler Ed., Berlin:Sringer-Verlag (LNCS Vol. 301), 1988, pp. 131-143.

[8] P.A. Devijver, "Real-time modelling of image sequence," *submitted to IEEE Transactions on PAMI*, 1990.

[9] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*, pp. 21-741,1984.

[10] X. Guyon et J. Yao, "Analyse discriminante contextuelle," Fifth International Symposium of Data Analysis and Informatics, (INRIA, B.P.105,78153 Le Chesnay Cedex), 1987.

[11] N. Marhic and W. Pieczynski, "Estimation of mixture and unsupervised segmentation of images," *Proceedings of IGARSS*, Helsinki, Finlande, 1991.

[12] J. Marroquin, S. Mitter and T. Poggio, " Probabilistic solution of ill-posed problems in computational vision," *Journal of the American Statistical Association, 82*, pp. 76-89, 1987.

[13] P. Masson and W. Pieczynski, "Segmentation of SPOT images by contextual SEM," *Proceedings of EUSIPCO*, Barcelone, Espagne, Sept. 1990.

[14] W. Pieczynski, "Missing data and iterative conditional estimation," Rapport technique, No.134, LSTA, Université de Paris 6, Fév. 1991.

[15] R.A. Redner and H.F. Walker, "Mixture densities, maximum likehood and the EM algorithm," *Society for Industrial and Applied Mathematics Review, Vol. 26*, No.2, pp. 195-239, Apr. 1984.