

Sur les segmentations statistiques non supervisées dans le contexte des mélanges généralisés

Laurent FOUQUE, Abdel-Nasser TEBBACHE, Wojciech PIECZYNSKI

Institut National de Télécommunications
Département Signal et Image
9 Rue Charles Fourier 91000 EVRY-France
laurent.fouque@sim.int-evry.fr

Résumé – Notre étude concerne le choix entre deux familles de méthodes statistiques non supervisées de segmentation d’images dans un contexte de mélange de lois généralisé. La première famille est la famille utilisant les champs de Markov cachés. La seconde, plus récente, est fondée sur l’utilisation des chaînes de Markov cachées, formées par la transformation de l’ensemble des pixels en un processus monodimensionnel par une transformation de type Peano. Dans les deux cas, nous sommes confrontés au problème de l’estimation d’un mélange de lois. Dans un cadre de mélange généralisé, la forme du bruit pour chaque classe est inconnue. Cependant, nous supposons qu’elle appartient à un ensemble de lois paramétrées. Nous utilisons une méthode générale d’estimation (ICE-GEMI) pour laquelle nous étudions deux nouveaux critères permettant de déterminer la forme de la loi. Finalement, nous sommes en présence de différentes méthodes de segmentation non supervisées que nous étudions afin d’apporter des éléments de réponse quant au choix de la meilleure méthode de segmentation à utiliser en fonction de l’image à traiter.

Abstract – The subject of this paper is the choice between two families of unsupervised statistical methods for image segmentation in the context of generalized distribution mixtures. The first family is the method using hidden Markov fields. The second, which is more recent, is based on the use of hidden Markov chains, formed by the transformation of the set of pixels into a one-dimensional process via a Peano type scan. In both cases, we have to deal with the problem of estimating a distribution mixture. In a generalized mixture context, the nature of the noise for each class is unknown. However, we suppose that it belongs to a set of families defined by parameters. We use a general estimation method (ICE-GEMI) for which we propose two new criteria that enable us to determine the form of the distribution. In this way, different unsupervised segmentation methods are available. We have analyzed these in order to provide answers concerning the best segmentation method to use according to the image being considered.

1 Introduction

Notre travail concerne le choix entre deux types de méthodes statistiques non supervisées de segmentation d’images, dans le contexte des mélanges généralisés. La première famille des méthodes est la famille utilisant les champs de Markov cachés [4][8]. La deuxième famille, plus récente, est fondée sur les chaînes de Markov cachées, formées par la transformation de l’ensemble des pixels en un processus monodimensionnel par une transformation de type Peano [1]. Ces deux types de familles ont déjà été comparés dans le cas classique de bruit gaussien et il s’est avéré que les méthodes fondées sur les chaînes, plus rapides, peuvent être compétitives [9]. Notre travail consiste à prolonger l’étude présentée dans [9] au cas des mélanges généralisés. Nous sommes en face du problème de l’estimation de mélange généralisé lorsque la forme du bruit pour chaque classe n’est pas connue. A titre d’exemple, considérons le cas de deux classes où le bruitage de chacune d’entre elle peut être gaussien ou exponentiel. Il y a alors quatre possibilités de mélanges classiques, et on doit en plus de l’estimation des paramètres, déterminer dans lequel de ces quatre cas on se trouve. Une famille générale de méthodes d’estimation de tels mélanges (ap-

pelée ICE-GEMI) a été proposée dans [5]. Un des éléments clé de ces méthodes est le suivant : en supposant qu’un échantillon est produit par une seule loi paramétrique, déterminer la forme de cette loi parmi un ensemble de formes admissibles. La méthode utilisée pour résoudre ce problème dans le contexte des exemples proposés dans [5] (lesquels concernent les chaînes de Markov) est la minimisation de la distance de Kolmogorov. Nous étudions dans ce travail deux autres solutions pour ce problème particulier : une fondée sur l’histogramme, l’autre fondée sur l’estimation des densités par noyaux de convolutions. Finalement, nous sommes en présence de six méthodes de segmentation non supervisées différentes : trois correspondant aux trois manières (Kolmogorov, histogramme, noyaux) de traiter le problème particulier mentionné ci-dessus dans le contexte des chaînes de Markov, et les trois autres reprennent ces mêmes procédures dans le contexte des champs de Markov. L’efficacité de ces méthodes variant en fonctions de l’image, le problème du choix de la méthode peut être important. Dans cet article, nous étudions les six méthodes de segmentations non supervisées afin de choisir la méthode la mieux adaptée en fonction de l’image traitée.

L’organisation de l’article est la suivante. La deuxième

partie rappelle brièvement les notions de modélisations par champs et chaînes de Markov cachés ainsi que les algorithmes de segmentations correspondants. Dans la troisième partie, nous abordons le problème de l'estimation des paramètres définissant notre mélange de lois. Quelques résultats sont donnés dans la quatrième partie. La cinquième partie conclut notre article.

2 Modélisation Probabiliste et Classification bayésienne

2.1 Modèles de Markov cachés

Soit S , $card(S) = N$, l'ensemble des pixels de l'image. Nous considérons deux processus aléatoires $\mathbf{X} = (X_s)_{s \in S}$ et $\mathbf{Y} = (Y_s)_{s \in S}$, \mathbf{X} modélisant l'image des classes inobservables, \mathbf{Y} l'image des observations. Chaque X_s prend ses valeurs dans l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_k\}$ et nous supposons que les Y_s sont à valeurs dans R . Nous notons $\mathbf{x} = (x_s)_{s \in S}$ et $\mathbf{y} = (y_s)_{s \in S}$ des réalisations des processus \mathbf{X} et \mathbf{Y} . Dans le cadre des champs de Markov, la loi de \mathbf{X} est donnée par :

$$P[\mathbf{X} = \mathbf{x}] = \beta e^{-U_\alpha(\mathbf{x})} \quad (1)$$

où l'énergie $U_\alpha(\mathbf{x})$ sera dans cet article donnée par

$$U_\alpha(\mathbf{x}) = \sum_{(s,t) \text{ voisins}} \Phi_\alpha(x_s, x_t) \quad (2)$$

et

$$\begin{aligned} \phi_\alpha(\omega_i, \omega_i) &= -\alpha \\ \phi_\alpha(\omega_i, \omega_j) &= \alpha \end{aligned} \quad (3)$$

Dans le cadre des chaînes de Markov, la loi de \mathbf{X} s'écrit au moyen d'une probabilité initiale $\pi_i = P(X_1 = \omega_i)$, $1 \leq i \leq k$ et d'une matrice de transition homogène t définie par $(t_{ij}) = P(X_{n+1} = \omega_j | X_n = \omega_i)$. La loi de \mathbf{X} est donnée par :

$$P[\mathbf{X} = \mathbf{x}] = \pi_{x_1} * t_{x_1, x_2} * \dots * t_{x_{N-1}, x_N}. \quad (4)$$

Plusieurs algorithmes de segmentation comme le MAP [4] et le MPM [8] peuvent être utilisés. Dans la suite, nous adoptons le MPM. Le principe est le suivant :

$$\hat{x}_n = \arg \max_{i=1, \dots, k} P[X_n = \omega_i | \mathbf{Y} = \mathbf{y}] \quad (5)$$

3 Estimation des Paramètres

3.1 Paramètres a priori

Dans le cas où \mathbf{X} est un champ de Markov, plusieurs estimateurs de α obtenus à partir de $\mathbf{Y} = \mathbf{y}$ ont été proposés [2] [10] [6]. Dans notre étude, nous effectuons de manière itérative des tirages selon la loi a posteriori au cours desquels nous estimons α à l'aide des probabilités $P(X_s = \omega_i | r)$, où r est le nombre de ω_i présents dans le voisinage. Ainsi, pour obtenir la valeur de notre estimation, on effectue la moyenne des 5 valeurs obtenues (pour

$r = 0, \dots, 4$).

Dans le contexte des chaînes de Markov, l'estimation des paramètres a priori est effectuée de manière itérative en calculant à chaque itération de nouveaux estimés à partir de la loi a posteriori [5].

3.2 Paramètres définissant les lois conditionnelles

Pour l'estimation des lois conditionnelles, nous nous plaçons dans le cadre général présenté dans [5]. Nous devons résoudre le problème de l'estimation d'un mélange de loi. La forme de chaque loi conditionnelle est supposée appartenir à un ensemble de formes paramétrées ψ (gaussienne, beta...). La méthode d'estimation ICE-GEMI est une méthode itérative permettant de choisir la forme de chaque densité conditionnelle parmi toutes les formes présentes dans ψ et d'estimer les paramètres correspondants. A l'itération p , l'algorithme se déroule de la manière suivante :

- Simuler une réalisation \mathbf{x}^p de \mathbf{X} selon sa loi a posteriori calculée avec les paramètres estimés à l'itération $p - 1$
- Pour chaque classe w_i , $1 \leq i \leq k$, on construit les ensembles $S_i^p = \{n | x_n^p = \omega_i\}$. Soit $y_i^p = (y_n)_{n \in S_i^p}$.
- Pour chaque classe, on décide de la forme de la densité conditionnelle correspondante au moyen d'un critère C calculé à partir des $y_i^p = (y_n)_{n \in S_i^p}$, puis on estime ses paramètres.

Nous étudions dans la suite trois critères de choix C différents dans le but de comparer leur performances. Le premier, proposé dans un contexte de chaînes de Markov [5], utilise la fonction de répartition empirique F_n des observations. Nous proposons deux autres critères, l'un basé sur l'histogramme empirique H_{emp} , l'autre basé sur l'estimateur à noyau de convolution f_{emp} de la densité. C est caractérisé par la minimisation d'une distance entre la fonction empirique considérée (densité ou fonction de répartition) et son équivalent théorique calculé avec les paramètres estimés.

3.2.1 Utilisation de l'histogramme empirique

Soit f une densité de probabilité sur un intervalle I . Soient Y_1, \dots, Y_n des variables aléatoires indépendantes, de densité commune f . L'histogramme empirique, souvent utilisé pour estimer f , est défini par le choix d'un point $x_0 \in I$ et d'une largeur de pas h . Soit N_j le nombre de réalisations Y_j tombant dans l'intervalle j défini par $[x_0 + jh, x_0 + (j+1)h[$. Sur cet intervalle, nous donnons à l'histogramme H_n le poids $\frac{N_j}{nh}$. Celui-ci nous permet d'obtenir sous l'histogramme une aire égale à 1 (identique aux densités de probabilités). Dans le cadre de nos simulations, nous prendrons $x_0 = \min(y_1, \dots, y_n)$. Nous fixons le pas à

$$h = \frac{\max(y_1, \dots, y_n) - \min(y_1, \dots, y_n)}{Nint} \quad (6)$$

avec $Nint$ le nombre d'intervalles de l'histogramme. Nous prenons $Nint = \lfloor \frac{\log n}{\log 2} \rfloor + 1$ [7]. Pour définir la forme de

la densité à retenir, nous utilisons la distance définie par

$$d = \left(\int_I (H_n(x) - f(x))^2 dx \right)^{\frac{1}{2}} \quad (7)$$

mieux adaptée que la distance $d' = \sup_{x \in I} (|H_n(x) - f(x)|)$ sensible à des écarts ponctuels importants entre la densité et l'histogramme.

3.2.2 Utilisation de l'estimateur à noyau de la densité

L'estimateur à noyau de convolution est couramment utilisé pour l'estimation de la densité de probabilité. A partir d'un échantillon (y_1, \dots, y_n) de taille n , l'estimateur à noyau de la densité de probabilité f est donné par

$$f_n(y) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{y - y_i}{h}\right) \quad (8)$$

où K est un noyau, qui est une application de R dans R , bornée, intégrable par rapport à la mesure de Lebesgue et d'intégrale 1. Nous considérons le noyau gaussien défini par $K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$. Le paramètre h est un nombre réel positif dépendant de la taille n de l'échantillon. La détermination analytique d'une valeur optimale de h , cruciale pour la qualité de l'estimation, fait intervenir la vraie densité de probabilité qui nous est inconnue. Plusieurs approches permettent d'estimer cette valeurs [11]. Nous prendrons comme valeur de h celle proposée dans [3] $h = s_n n^{-\frac{1}{5}}$ avec s_n l'écart-type empirique des observations. La distance retenue pour le choix de la forme de la densité sera encore la distance définie par (7).

3.2.3 Utilisation de la fonction de répartition empirique

L'estimateur utilisé pour la fonction de répartition F est la fonction de répartition empirique donnée par

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, y]}(y_i) \quad (9)$$

Le choix sera effectué en considérant la distance de Kolmogorov définie par $d = \sup_{y \in R} |F_n(y) - F(y)|$.

4 Simulations

Nous avons effectué les simulations sur différentes images modélisées par un mélange de lois généralisé. Nous présentons dans les tableaux suivants quelques résultats que nous obtenons. Nous utilisons une image relativement homogène ("cible") puis une image très inhomogène ("ville"). Ces différentes images sont chacune constituées de 2 classes. L'image "cible" sera bruitée par une loi normale $\mathcal{N}(0;1)$ pour la classe 1 et par une loi exponentielle de paramètres $(\lambda = 1.5; u = 0)$ pour la classe 2 ($f(x) = \lambda \exp(-\lambda(x - u))$ pour $x \geq 0, 0$ sinon). L'image "ville" est bruitée par une loi exponentielle de paramètres $(1; -0.5)$ pour la classe 1 et par une loi Beta de paramètres $(3; 1.5; 0; 1)$ pour la classe 2. Nous obtenons un taux d'erreur de classification aveugle égale à 32% pour la première

| Champs - cible | H_{emp} | f_{emp} | F_n |
|----------------|-----------|-----------|-------|
| Lois | B - B | B - B | B - E |
| Erreur MPM | 7.7 | 8.1 | 6.1 |

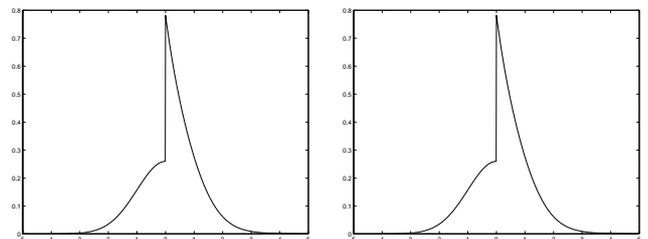
| Chaînes - cible | H_{emp} | f_{emp} | F_n |
|-----------------|-----------|-----------|-------|
| Lois | N - B | N - B | B - B |
| Erreur MPM | 14 | 14.2 | 10.3 |

| Champs - ville | H_{emp} | f_{emp} | F_n |
|----------------|-----------|-----------|-------|
| Lois | B - B | B - B | B - E |
| Erreur MPM | 19 | 18.2 | 14.5 |

| Chaînes - ville | H_{emp} | f_{emp} | F_n |
|-----------------|-----------|-----------|-------|
| Lois | B - E | B - E | B - E |
| Erreur MPM | 11.4 | 11.2 | 11.6 |

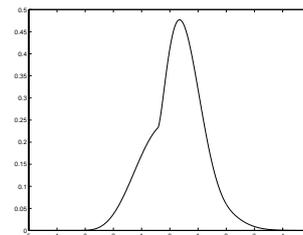
TAB. 1 – Lois reconnues et taux d'erreur MPM

image et 21% pour la seconde. Les résultats obtenus sont présentés dans le tableau 1. Les segmentations sont présentées dans la figure 2.



(a) mélange théorique

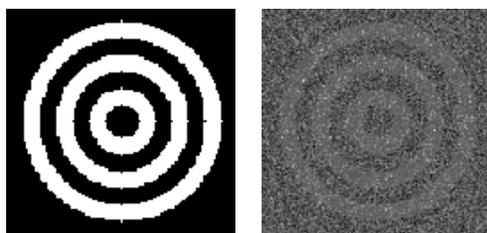
(b) estimé cible (champs et répartition)



(c) estimé cible (Chaîne et répartition)

FIG. 1 – Mélange théorique et exemples d'estimés ("cible")

Nous pouvons remarquer que l'utilisation des champs de Markov donne de meilleurs résultats de segmentation pour le cas de l'image cible, relativement homogène, mais que les résultats s'inversent pour le cas de l'image ville, fortement inhomogène donc mieux modélisée par une chaîne de Markov via le parcours de Hilbert. Par ailleurs, nous notons une similarité dans le comportement des méthodes utilisant l'histogramme et le noyau, ce qui est normale, l'histogramme n'étant qu'un cas simple et particulier de noyau. Enfin, l'utilisation de la fonction de répartition s'avère plus efficace que l'utilisation de l'histogramme et du noyau gaussien.



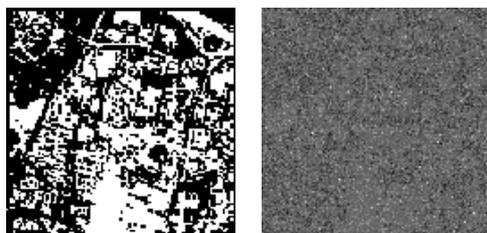
(a) image cible

(b) image Bruité



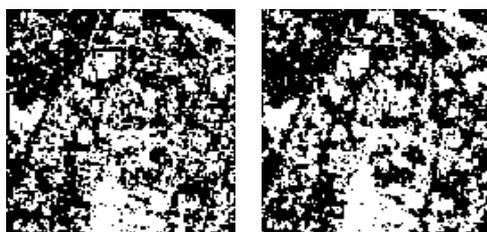
(c) image segmenté (chaîne et répartition)

(d) image segmenté (champs et répartition)



(e) image ville

(f) image Bruité



(g) image segmenté (chaîne et répartition)

(h) image segmenté (champs et répartition)

FIG. 2 – Résultats des Segmentations

5 Conclusion

Nous avons traité dans cet article du problème de la segmentation statistique non supervisée (dans un contexte de mélange généralisé) d'images. Nous avons proposé différentes variantes originales de la récente méthode générale

d'estimation ICE-GEMI [5] et comparé leurs efficacités sur des images de synthèse. D'une part, nous avons étudié deux sous-méthodes générales de ICE-GEMI : une fondée sur les chaînes de Markov cachées et une autre sur les champs de Markov cachés. Il apparaît qu'aucune de ces sous-méthodes n'est uniformément plus efficace que l'autre. D'autre part, nous avons testé trois manières de traiter un point particulier de ICE-GEMI consistant en choix, pour chaque classe et à chaque itération, d'une forme de densité de probabilité dans un ensemble de formes admissibles. Le choix fondé sur le calcul des fonctions de répartitions empiriques et la minimisation de la distance de Kolmogorov semble préférable aux méthodes concurrentes.

Références

- [1] Benmiloud, B. et Pieczynski, W. "Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images". *Traitement du Signal*, Vol 12, n. 5, pp 433-454, 1995.
- [2] Besag, J. "On the statistical analysis of dirty pictures". *Journal of the royal statistical society, B*, Vol 48, pp 259-302, 1986.
- [3] Bosq, D. "Non parametric statistics for stochastic processes". *Lecture Notes in Statistics*, n. 110, Springer Verlag, 1996.
- [4] Geman, S. and Geman, S. "Stochastic relaxation, Gibbs distributions, and the bayesian restoration of Images". *IEEE Trans. on PAMI*, vol 6, n. 6, pp 721-741, 1984.
- [5] Giordana, N. and Pieczynski, W. "Estimation of generalized multisensor hidden markov chains and unsupervised image segmentation". *IEEE Trans. on PAMI*, vol 19, n. 5, pp 465-475, 1997.
- [6] Guyon, X. "Champs aléatoires sur un réseau", *Collection Techniques Stochastiques*, Masson, Paris, 1993.
- [7] Law, A. M. and Kelton, W. D. "Simulation modeling and analysis", McGraw-Hill International editions, 1991.
- [8] Marroquin, J. L. ;Mittle, S. and Poggio, T. "Probabiliste solution of ill-posed problems in computational vision". *Journal of the American Statistical Association*, vol 82, pp 76-89, 1987.
- [9] Salzenstein, F. and Pieczynski, W. "Sur le Choix de Méthode de segmentation statistique d'images". *Traitement du signal*, vol 15, n. 2, pp 119-128, 1998.
- [10] Younes, L. "Parametric inference for imperfectly observed Gibbsian fields". *Prob.Theory Related Fields*, vol 82, pp 625-645, 1989.
- [11] Silverman, B. W. "Density estimation". Chap 3, Chapman and Hall, New York, 1986.