Estimation Conditionnelle Itérative dans les chaînes de Markov cachées et segmentation statistique non supervisée d'images

B. Benmiloud A. Peng W. Pieczynski

Groupe Image Département Système et Réseaux Institut National des Télécommunications.
9, rue Charles Fourier, 91011 Evry Cédex, France

RÉSUMÉ

Nous proposons dans cet article une méthode originale d'estimation des paramètres dans les chaînes de Markov cachées avec application à la segmentation statistique non supervisée d'images. Notre algorithme est fondé sur une méthode récente générale d'estimation dans le cas des données incomplétes dite "Estimation Conditionnelle Itérative" (ECI). Nous montrons la supériorité de l'efficacité de la segmentation ainsi obtenue sur celle de la segmentation contextuelle.

ABSTRACT

In this paper we propose a new method of hidden Markov chains parameter estimation with application to unsupervised statistical image segmentation. Our algorithm is based on a recent general method of parameter estimation in the case of incomplet data, so called "Iterative Conditonal Estimation" (ICE). We show the better effectiveness of the segmentation method so obtained with respect to the contextual segmentation one.

1. Introduction

Les approches statistiques de l'important problème de la segmentation d'images ont été développées par de nombreux auteurs ces dernières années. Les méthodes utilisées sont basées sur la modélisation d'images par des composantes markoviennes. Le modèle proposé dans cet article est fondé sur la technique des chaînes de Markov cachées. Cette modélisation est largement exploitée en traitement de la parole, mais elle peut être transposée au bénéfice de l'image. L'image est modélisée par un couple $(X=(X_{s\in S}), Y=(Y_{s\in S}))$ de champs aléatoires où $S=\{1,...,n\}$ représente l'ensemble des pixels. Chaque X_s prend ses valeurs dans l'ensemble discret des classes $\Omega = \{\omega_1,..,\omega_k\}$, et Y_s est une variable aléatoire réelle. La la réalisation de X n'est pas accessible à l'observation. Par contre, ce qui est observé est la version bruitée du champ X; c'est le modèle de Markov cachée. Le but est d'estimer la réalisation X à partir de la réalisation du champ Y. La loi du couple (X,Y) est définie par la loi de X, dite "a priori", et les densités conditionnelles de Y sachant X. Le problème crucial qui se pose est l'étape d'estimation des paramètres, ce qui permet d'aboutir à la segmentation non supervisée. Nous proposons une nouvelle méthode d'estimation de tous les paramètres de la chaîne de Markov cachée obtenue à partir d'une méthode générale récénte dite "Estimation Conditionnelle Itérative" [2]. Cette nouvelle procédure s'avère compétitive vis à vis des algorithmes EM [1, 5] et SEM [4]. Notre objectif est donc d'étudier le comportement de la méthode ECI dans le cadre des chaînes de Markov cachées et de le comparer à une famille des méthodes contextuelles où le contexte utilisé se réduit à un seul voisin. Nous présentons plus précisément des résultats de la segmentation non supervisée des images binaires respectivement avec les deux méthodes. La segmentation utilisée dans le cadre des chaînes de Markov cachées est l'algorithme MPM: chaque pixel est classé à partir de toute l'image. Méthode globale de segmentation, elle est basée sur l'algorithme "backward - forward" [1]. Dans le contexte étudié l'image est considérée comme une chaîne de Markov, or ce passage d'un modèle bidimensionnel à un modèle unidimensionnel entraîne le problème suivant : le "passé" d'un pixel ne correspond pas exactement à son contexte spatial, d'où l'utilisation d'un parcours autre que le parcours classique. Les courbes de Hilbert-Péano [6], par construction, ne possédent aucun point double et peuvent remplir tout l'espace plan, tout en permettant de tenir compte au maximum du contexte spatial de chaque pixel. L'utilisation de cette famille de courbes améliore considérablement le taux de restauration des images [3]. La chaîne de Markov est un processus stochastique causal, la loi de probabilité sur un pixel dépend seulement des états précédents et non des états suivants, tandis qu'en présence d'un champ markovien, la loi de probabilité sur un site dépend des "voisins spatiaux". Cette modélisation par les chaînes de Markov nous offre la possibilité d'avoir des algorithmes rapides d'estimation pour une étude ultérieure de la segmentation des séquences d'images.

2. Le modèle des chaînes de Markov cachées

X est une variable aléatoire dont la loi de distribution est une chaîne de Markov homogène, stationnaire, d'ordre 1, à valeurs dans l'ensemble des classes $\Omega = \{\omega_1,..,\omega_k\}$. Sa loi est alors donnée par la matrice de transition :

$$a_{ij} = \frac{P(X_m = \omega_i\, ,\, X_{m+1} = \omega_j)}{\sum_{j=1}^{k} P(X_m = \omega_i\, ,\, X_{m+1} = \omega_j)}$$

Notons que les lois marginales

$$q_i = P(X_m = \omega_i)$$

sont données par la matrice de transition et ne dépendent pas de m.

La probabilité d'une réalisation $X = x = (\omega_{i_1},...,\omega_{i_n})$ s'écrit :

$$f(x) = q_{i_1}\, a_{i_1 i_2} .... a_{i_{n-1} i_n}$$

f étant la densité par rapport à la mesure de dénombrement. Les variables aléatoires (Y_s) sont supposées indépendantes conditionnellement à X, la loi de chaque Y_s conditionnelle à X est de plus supposée être égale à sa loi conditionnelle à X_s. La loi de Y conditionnelle à X est donc déterminée par les k lois de Y_s conditionnelles à $X_s = \omega_1,...,\omega_k$. Notons $f_1,...,f_k$, les densités de ces lois par rapport à la mesure de Lebesgue et $\beta_1,..., \beta_k$ les paramètres définissant ces densités. La loi de distribution de (X,Y) en (x,y) est la suivante :

$$h(x,y) = q_{i_1} f_{i_1}(y_1) a_{i_1 i_2} f_{i_2}(y_2) ... \, a_{i_{n-1}} a_{i_n} f_{i_n}(y_n)$$

3. Principe général de ECI

La loi du couple (X, Y) est donc définie par : la loi de X, appelée loi a priori, et les lois de Y conditionnelles à X déterminées respectivement par l'ensemble des paramètres α et β. Ces derniers sont inconnus. Le problème consiste à estimer $\theta = (\alpha, \beta)$ à partir de l'observation du champ Y. L'algorithme ECI est une procédure récente d'estimation des paramètres dans le cas des données cachées [2]. Nous cherchons un estimateur de θ à partir de Y, la seule variable aléatoire observable. Considérons $\hat{\theta} = \hat{\theta}$ (X, Y), un estimateur de θ à partir de (X,Y). Le principe de ECI est d'approcher l'estimateur $\hat{\theta}$ par une fonction de Y. La meilleure approximation, au sens de l'erreur quadratique moyenne, est l'espérance conditionnelle. L'espérance conditionnelle de $\hat{\theta}$ par rapport à Y, notée $E[\hat{\theta}/Y]$, dépend du paramètre θ. Pour calculer cette espérance, nous utilisons une démarche itérative qui consiste à calculer la valeur suivante en fonction de la valeur "courante" :

$$\theta_{q+1} = E_q[\hat{\theta}\ /Y]$$

sachant que E_q est l'espérance conditionnelle correspondante à θ_q. La procédure de l'algorithme ECI est la suivante :

- L'initialisation du paramètre θ_0
- θ_{q+1} est calculé à partir de θ_q et Y=y par :

$$\theta_{q+1} = E_q[\hat{\theta}\ /Y].$$

Dans notre cadre d'étude, il s'agit d'estimer les paramètres suivants :

- la probabilité a priori de X est donnée par $\alpha = (\alpha_{ij})$ défini par :

$$\alpha_{ij} = P[X_s = \omega_i, X_{s+1} = \omega_j]$$

et on a :

$$q_i = \sum_{j=1}^{k} \alpha_{ij}$$

si X était observable, on pourrait considérer l'estimateur suivant :

$$\widehat{\alpha_{ij}} = \frac{1}{n-1} \sum_{s=1}^{n-1} 1[X_s = \omega_i, X_{s+1} = \omega_j]$$

son espérance conditionnelle à l'itération q+1 est alors :

$$\alpha_{ij}^{q+1} = E_q\ [\widehat{\alpha_{ij}} / Y=y]$$

ce qui donne la matrice de transition correspondante à l'itération q+1 :

$$a_{ij}^{q+1} = \frac{\alpha_{ij}^{q+1}}{\sum_{j=1}^{k} \alpha_{ij}^{q+1}}$$

En appliquant la définition de l'espérance conditionnelle, on obtient ainsi une formule similaire à celle obtenue par la méthode développée par Baum [7]. Elle s'écrit en fonction de la probabilité conditionnelle de transition de la classe ω_i à celle de ω_j notée P_y :

$$a_{ij}^{q+1} = \frac{\sum_{s=1}^{n-1} P_y\ (X_s = \omega_i, X_{s+1} = \omega_j)}{\sum_{j=1}^{k}\sum_{s=1}^{n-1} P_y\ (X_s = \omega_i, X_{s+1} = \omega_j)}$$

Elle peut s'exprimer en fonction des probabilités forward et backward comme suit :

$$P_y(X_s = \omega_i, X_{s+1} = \omega_j) = F_s(\omega_i)B_{s+1}(\omega_j)a_{ij}^{q} f_j(y_{s+1})$$

d'où la matrice de transition :

$$a_{ij}^{q+1} = \frac{\sum_{s=1}^{n-1} F_s(\omega_i)B_{s+1}(\omega_j)a_{ij}^{q} f_j(y_{s+1})}{\sum_{s=1}^{n-1} F_s(\omega_i)B_s(\omega_i)}$$

Nous retrouvons la même formule que celle de l'algorithme EM [1] pour la réestimation des paramètres de transition. Pour le développement des calculs voir la référence [1 et 7]. Soit $F_s(\omega_j)$ la probabilité forward définie par :

$$F_s(\omega_j) = \frac{P(X_s=\omega_j, Y_1, \dots, Y_s)}{P(Y_1, \dots, Y_s)}$$

et $B_s(\omega_j)$ la probabilité backward définie par :

$$B_s(\omega_j) = \frac{P(Y_{s+1}, \dots, Y_n / X_s=\omega_j)}{P(Y_{s+1}, \dots, Y_n / Y_1, \dots, Y_s)}$$

- Pour une distribution Gaussienne que nous considérons dans la suite β_i est donné par la moyenne m_i et la variance σ_i^2.

En ce qui concerne ce paramètre β, le calcul de l'espérance conditionnelle de son estimateur par rapport à Y=y n'est pas envisageable. Or la loi des grands nombres nous permet d'estimer β^{q+1} en faisant des tirages stochastiques. Ces tirages se font selon les lois a posteriori dans l'ensemble Ω. On obtient une partition de $(y_1, \dots y_n)$ notée : Q_j^q, avec j désignant la classe de ω_j. Les paramètres du bruit sont alors calculées par les formules empiriques données par :

- La moyenne de la classe ω_j est :

$$m_j^{q+1} = \frac{\sum_{s=1}^{n} y_{s,j}^{q}}{card(Q_j^q)}$$

- La variance correspondante est donnée par:

$$\sigma_j^{2q+1} = \frac{\sum_{s=1}^{n} (y_{s,j}^{q} - m_j^{q+1})}{card(Q_j^q)}$$

Remarquons que le paramètre a priori est estimé par une méthode déterministe, tandis que les paramètres du bruit sont obtenus par des tirages stochastiques selon la probabilité a posteriori, en simulant des réalisations à chaque itération. Cette méthode dépend peu de l'initialisation contrairement aux méthodes déterministes du type EM.

4. Résultats de la segmentation non supervisée d'images et discussions

Nous considérons une image binaire im (fig.1), présentant des anneaux, bruitée avec deux types de bruit Gaussien indépendant suivants : imbr1 (fig.2) ($m_0 = 110$, $m_1 = 140$ et $\sigma_0^2 = \sigma_1^2 = 900$) et imbr2 (fig.3) ($m_0 = m_1 = 125$ et $\sigma_0^2 = 225$, $\sigma_1^2 = 900$). Les paramètres de bruits et les probabilités a priori sont estimés par l'algorithme ECI, puis utilisés pour la segmentation de l'image par l'algorithme MPM. Les images obtenues lors de cette segmentation sont imbr11 (fig.4) et imbr21 (fig.5), cette méthode est comparée avec les images obtenues par la segmentation contextuelle non supervisée [5], pour chaque type de bruits. Elles correspondent respectivement aux images imbr12 (fig.6) et imbr22 (fig.7). La méthode MPM couplée à l'estimateur ECI donne des résultats nettement meilleurs que les méthodes contextuelles. Notons τ, le taux d'erreur de classification en pourcentage.

Fig.1

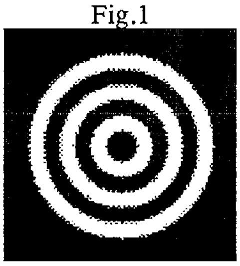


Fig.2

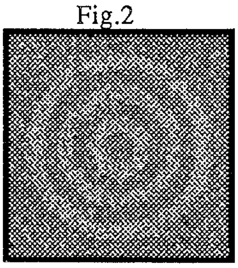


Fig.3

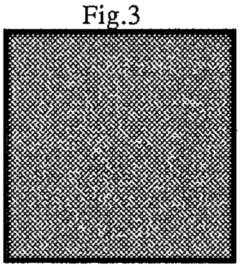
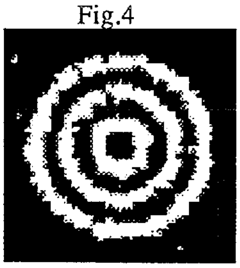
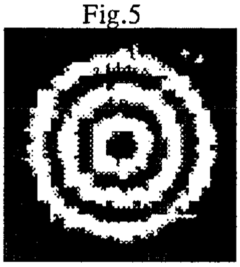


Fig.4



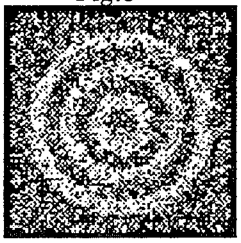
$\tau = 7.51$ %

Fig.5



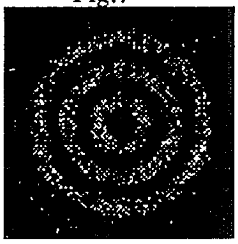
$\tau = 8.38$ %

Fig.6



τ=33.60 %

Fig.7



$\tau = 26.32$%

Les résultats obtenus montrent l'intérêt de l'utilisation du modèle des chaînes de Markov cachées dans la segmentation non supervisée d'images. L'erreur de classification atteint dans certains cas le taux d'erreur théorique, i.e. calculé sur la base des vrais paramètres, ce qui montre le bon comportement de ECI dans ce contexte. L'étape d'apprentissage dans le cas non supervisé est très importante dans la segmentation. Une étude comparative des algorithmes EM et SEM et ECI, dans le cadre de la segmentation contextuelle, a été faite [5]. Le comportement de ces trois algorithmes semble, dans le cadre considéré, équivalent.

5. Références

[1] P. A. Devijver, M. Dekesel, Champs aléatoires de Pickard et modélisation d'images digitales. Traitement du signal, vol 5, n° 5, 1988.
[2] W. Pieczynski, Parameter estimation in the case of hidden data, 16 th biennial Symposium on Communication, Kingston, Canada, Mai 1992.
[3] M. Emsalem, H. Caillol, P. Olivié, G. Carnat, W. Pieczynski, Fast Unsupervised Statistical Image Segmentation Method, IGARSS 92, Houston,Texas, Mai 1992.
[4] G. Celeux, J. Diebolt, L'algorithme SEM: un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités, Revue de la Statistique Appliquée, Vol. 34, No 2, 1986.
[5] A. Peng, Segmentation statistique non supervisée d'images et détection de contours par filtrage, Thèse de Doctorat, UTC, Novembre 92.
[6] J. Quinqueton, Le concept de dimension intrinsèque en reconnaissance des formes, Thèse de Doctorat, UPMC, Février 81.
[7] R. Boite et M. Kunt, Traitement de la parole, Presses Polytechniques Romandes.