

## Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images

---

### *Parameter Estimation in Hidden Markov Chains and Segmentation of Images*

par **Btissam BENMILOU**, **Wojciech PIECZYNSKI**

*Département Signal et Image  
Institut National des Télécommunications  
9, rue Charles Fourier  
91011 Évry cedex*

#### Résumé

Notre étude traite de l'estimation des paramètres dans les chaînes de Markov cachées et de la segmentation statistique non supervisée d'images. Nous proposons deux algorithmes originaux d'estimation obtenus à partir des méthodes Iterative Conditional Estimation (ICE) et Stochastic Expectation Maximisation (SEM), notés MICE et MSEM respectivement, et montrons leur compétitivité vis-à-vis de l'algorithme Expectation Maximisation (EM) dans différents cas d'homogénéité et de bruitage des chaînes. L'étude du comportement des trois algorithmes de restauration non supervisée des chaînes obtenus par l'adjonction à la méthode Mode de la Marginale a Posteriori (MPM) des algorithmes EM, MICE, MSEM respectivement est ensuite proposée. La transformation des processus bi-dimensionnels en processus mono-dimensionnels par le parcours de Peano rend possible l'application de ces algorithmes au problème de la segmentation statistique non supervisée d'images. On obtient ainsi des méthodes plus rapides que celles utilisant des modélisations par champs de Markov cachés et nous montrons que la perte de l'efficacité est, en général, acceptable. La souplesse de notre modélisation permet par ailleurs la conception de nombreux algorithmes de segmentation statistique non supervisée spatio-temporelle d'images. Nous en proposons trois et présentons les résultats de leur application à la segmentation d'une séquence d'images réelles.

**Mots clés :** Chaînes de Markov cachées, estimation des paramètres, segmentation non supervisée, segmentation spatio-temporelle, expectation-maximisation, iterative conditional estimation, stochastic expectation-maximisation, backward-forward algorithm.

#### Abstract

*Our study deals with the parameter estimation problem of Hidden Markov Chain models and with unsupervised Bayesian image segmentation. We propose two new estimation algorithms obtained from Iterative Conditional Estimation (ICE) and Stochastic Expectation Maximisation (SEM), denoted by MICE and MSEM respectively, and show their competitiveness with respect to the Estimation Maximisation (EM) algorithm in different situations of chain homogeneity and noise. We then study three unsupervised chain restoration algorithms, obtained by adding EM, MICE and MSEM respectively to the Maximum Posterior Mode (MPM) restoration method. The transformation of bi-dimensional process to mono-dimensional ones using Peano curves makes possible the application of these three methods to the problem of unsupervised statistical image segmentation. Doing so, we obtain faster methods than those obtained by models using hidden Markov random fields and we show that the loss of effectiveness, due to the poorer adequacy of the model, is acceptable in general. On the other hand, the flexibility of our modeling allows the conception of numerous unsupervised spatio-temporal segmentation methods. We propose three of them and present results showing their application to the segmentation of a sequence of real images.*

**Key words :** Hidden Markov chains, parameter estimation, unsupervised segmentation, spatio-temporal segmentation, expectation-maximisation, iterative conditional estimation, stochastic expectation-maximisation, backward-forward algorithm.

---

## 1. Introduction

Les modélisations par chaînes de Markov cachées permettent d'aborder de nombreux problèmes, plus particulièrement en traitement de la parole ([9], [39], [40]) ou d'image ([5], [17], [18], [19], [37]). On considère que l'on a accès à une version bruitée du signal modélisé par une chaîne de Markov et le problème général est celui de l'estimation de la réalisation inobservable de

la chaîne. La modélisation générale permet l'application des techniques bayésiennes et les résultats obtenus sont, en général, tout à fait satisfaisants. En situations réelles les paramètres définissant le modèle sont inconnus dans de nombreux cas. On est ainsi confronté, avant même l'application d'une méthode de restauration donnée, au problème de l'estimation de tous les paramètres du modèle à partir du signal bruité, le seul disponible. On arrive au problème général, et difficile, de reconnaissance de mélange de lois de probabilité. Un certain nombre de techniques comme Ex-

pectation Maximisation (EM, [16]), Stochastic Expectation Maximisation (SEM, [14]), Iterative Conditional Estimation (ICE, [33], [34]), l'algorithme de Viterbi non supervisé ([2], [43]) a été proposé par les auteurs.

Nous poursuivons dans ce travail un double objectif. D'une part, nous proposons deux algorithmes originaux d'estimation de tous les paramètres de chaînes de Markov cachées et comparons leur comportement, dans diverses situations, avec celui de l'algorithme EM. Cet objectif est général et les résultats obtenus se situent en dehors des problèmes propres à l'imagerie. D'autre part, nous appliquons nos méthodes à la problématique de la segmentation statistique non supervisée d'images. Une étude comparative de trois méthodes de segmentation d'images fixes est suivie de la proposition des algorithmes originaux de segmentation non supervisée spatio-temporelle.

Nos algorithmes de segmentation non supervisée sont différents aussi bien de ceux utilisant les modélisations par champs de Markov cachés ([3], [8], [11], [12], [15], [21], [22], [23], [24], [26], [27], [28], [29], [34], [44], [45]) que ceux utilisant les approches locales ([12], [13], [22], [28], [30], [31], [32], [42], [44]). Nous proposons ainsi une troisième famille de méthodes dont une certaine souplesse, très utile en segmentation spatio-temporelle, est la principale caractéristique.

L'organisation de l'article est la suivante :

Les sections deux et trois sont consacrées au rappel de la modélisation et des algorithmes de restauration. Dans la section quatre nous rappelons les démarches générales des méthodes d'estimation des paramètres EM, ICE, SEM et proposons deux algorithmes originaux obtenus à partir de ICE et SEM (notés MICE, MSEM respectivement). L'étude du comportement des méthodes EM, MICE, MSEM est présentée dans la section cinq. Les trois algorithmes de restauration non supervisée des chaînes de Markov cachées obtenus par l'adjonction à la méthode de restauration MPM des algorithmes EM, MICE, MSEM respectivement sont étudiés dans la section six. La septième section est consacrée à la segmentation non supervisée des images fixes. Nous y proposons les résultats de diverses segmentations des images de synthèse. Le problème de la segmentation spatio-temporelle est abordé dans la huitième section. Nous proposons trois démarches originales et présentons les résultats de segmentation d'une séquence d'images réelles. La neuvième et dernière section comporte les conclusions.

## 2. Les chaînes de Markov cachées

Une chaîne de Markov cachée est un processus à temps discret doublement stochastique, ou encore composé de deux processus  $X = (X_n)_{n \in N}$  et  $Y = (Y_n)_{n \in N}$ . Le processus  $X$  est une chaîne de Markov et le processus  $Y$  est réel (chaque  $Y_n$  prend ses valeurs dans  $R$ ). L'appellation « cachée » signifie que les réalisations de  $X$  sont inobservables. Le problème général, modélisant de nombreux

problèmes réels, est alors celui de l'estimation de la réalisation de  $X$  à partir de la réalisation observée de  $Y$ , ou encore celui de la « restauration » de  $X$ . Le terme de restauration se justifie dans l'optique du vocabulaire du traitement du signal : en adoptant ce dernier, le processus  $Y$  peut être considéré comme une version « bruitée » du processus  $X$ . La notion du « bruit » est cependant à considérer dans un sens très général. En traitement de la parole un discours peut être considéré comme une suite de phonèmes : le  $n$ ème phonème est ainsi la réalisation de  $X_n$ . Le bruit modélise ici le fait qu'un phonème particulier est prononcé différemment selon les personnes : la mesure représentant une prononciation particulière est ainsi la réalisation de  $Y_n$ . L'objectif du modèle est ici de permettre la conception des méthodes de transcription des discours indépendantes du discours considéré (aspect aléatoire de  $X$ ), mais aussi indépendantes de la personne qui le prononce (aspect aléatoire de  $Y$  conditionnellement à  $X$ ). Considérons par ailleurs une image numérique satellite où de l'eau et de la forêt sont présentes. On souhaite établir une carte, ce qui revient à associer à chaque pixel un élément dans l'ensemble  $\Omega = \{\text{eau}, \text{forêt}\}$ . Sur le pixel  $n$  la réalisation invisible de  $X_n$  est ainsi « eau » ou « forêt » et la réalisation observée de  $Y_n$  est le niveau de gris (un nombre). Le bruit modélise ici la « variabilité naturelle » de l'eau et de la forêt : les deux classes ne produisent pas nécessairement deux mesures uniques. D'autres bruits (transmission, acquisition, ...) peuvent éventuellement s'ajouter au bruit « variabilité naturelle ». Le bruit global est donc modélisé par les lois de  $Y_n$  conditionnelles à  $X_n$ .

Nous rappelons dans ce paragraphe les définitions relatives aux diverses lois de probabilité liées aux processus  $X$  et  $Y$  ainsi que les modes de leurs calculs, notions nécessaires pour aborder les problèmes de la restauration et de l'estimation des paramètres du modèle.

### 2.1. LOI DE $X$

Une suite de variables aléatoires  $(X_n)_{n \in N^*}$  prenant leurs valeurs dans l'ensemble des états (ou classes)  $\Omega = \{\omega_1, \dots, \omega_k\}$  est une chaîne de Markov si elle vérifie pour tout  $n \geq 1$  :

$$P[X_n = x_n / X_1 = x_1, \dots, X_{n-1} = x_{n-1}] = P[X_n = x_n / X_{n-1} = x_{n-1}] \quad (1)$$

La loi du processus  $X = (X_n)_{n \in N^*}$  est alors déterminée par la loi  $\pi$  de  $X_1$ , dite loi initiale, et la suite de matrices de transition  $a_{ij}^n = P[X_{n+1} = \omega_j / X_n = \omega_i]$ . Nous supposons dans la suite que

$$c_{ij} = P[X_n = \omega_i, X_{n+1} = \omega_j] \quad (2)$$

ne dépendent pas de  $n$ . La loi initiale est alors donnée par

$$\pi_i = P[X_1 = \omega_i] = \sum_{1 \leq j \leq k} c_{ij} \quad (3)$$

et la matrice de transition  $A = [a_{ij}]$ , également indépendante de  $n$ , donnée par

$$a_{ij} = P[X_{n+1} = \omega_j / X_n = \omega_i] = \frac{c_{ij}}{\sum_{1 \leq j \leq k} c_{ij}} \quad (4)$$

Ainsi, par le théorème de Kolmogorov, les  $k^2$  paramètres  $(c_{ij})_{1 \leq i \leq k, 1 \leq j \leq k}$  déterminent entièrement la loi du processus  $X$ . La loi de probabilité du vecteur aléatoire  $X^n = (X_1, \dots, X_n)$  s'écrit :

$$P[X_1 = \omega_{i_1}, X_2 = \omega_{i_2}, \dots, X_n = \omega_{i_n}] = \pi_{i_1} a_{i_1 i_2} \dots a_{i_{n-1} i_n} \quad (5)$$

## 2.2. LOI DE $(X, Y)$

La loi de  $X$  étant définie ci-dessus il reste à définir les lois de  $Y$  conditionnelles à  $X$ . Nous supposons que les variables aléatoires  $(Y_n)$  sont indépendantes conditionnellement à  $X$ . De plus, la loi de chaque  $Y_n$  conditionnelle à  $X$  est égale à sa loi conditionnelle à  $X_n$ . Nous noterons  $f_1, f_2, \dots, f_k$  les densités des lois de  $Y_n$  conditionnelles à  $X_n = \omega_1, X_n = \omega_2, \dots, X_n = \omega_k$  respectivement. Ainsi les densités  $f_1, f_2, \dots, f_k$  définissent toutes les lois de  $Y$  conditionnelles à  $X$ .

Dans la suite  $n$  sera fixé et nous noterons  $X = (X_1, \dots, X_n)$ ,  $Y = (Y_1, \dots, Y_n)$  les suites finies de variables aléatoires et  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$  respectivement leurs réalisations.  $f_x$  étant la densité de la loi de  $Y$  conditionnelle à  $X = x$ , la loi de  $(X, Y)$  est définie par :

$$h(x, y) = P[X = x] f_x(y) \quad (6)$$

qui est sa densité par rapport au produit de la mesure de dénombrement sur  $\Omega^n$  par la mesure de Lebesgue sur  $R^n$ .

Nous avons vu que la probabilité de  $X = x = (\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_n})$  s'écrit :

$$P[X = x] = \pi_{i_1} a_{i_1 i_2} \dots a_{i_{n-1} i_n} \quad (7)$$

Étant donné l'indépendance des  $Y_m$  conditionnellement à  $X$ , on a la factorisation suivante :

$$f_x(y) = \prod_{m=1}^n f_{i_m}(y_m) \quad (8)$$

finaleme $n$ t  $h$  s'écrit :

$$h(x, y) = \pi_{i_1} f_{\omega_{i_1}}(y_1) a_{i_1 i_2} f_{\omega_{i_2}}(y_2) \dots a_{i_{n-1} i_n} f_{\omega_{i_n}}(y_n) \quad (9)$$

## 2.3. LOI DE $X$ A POSTERIORI

La connaissance de la loi de  $X$  a posteriori, i.e. conditionnelle à  $Y = y$ , est cruciale pour envisager la recherche des solutions du

problème posé, à savoir l'estimation de la réalisation cachée de  $X$  à partir de la réalisation observée  $Y = y$ . Cette loi contient en effet toute l'information sur  $X$  contenue dans l'observation  $Y = y$ . Elle est donnée par :

$$P^{Y=y}[X = x] = P[X = x / Y = y] = \frac{h(x, y)}{\sum_{x \in \Omega^n} h(x, y)} \quad (10)$$

On montre que la distribution de  $X$  a posteriori est celle d'une chaîne de Markov (non stationnaire). Plus précisément, les matrices de transition de cette distribution sont :

$$\begin{aligned} a_{ij}^{m+1} &= P^{Y=y}[X_{m+1} = \omega_j / X_m = \omega_i] \\ &= \frac{a_{ij} f_j(y_{m+1}) \beta_{m+1}(j)}{\sum_{j=1}^k a_{ij} f_j(y_{m+1}) \beta_{m+1}(j)} \end{aligned} \quad (11)$$

où  $\beta_j(y_{m+1})$ , dite probabilité « backward », est définie dans le paragraphe suivant.

Cela implique que la simulation des réalisations de  $X$  selon sa distribution a posteriori peut être faite directement, ce qui constitue un avantage important, lorsqu'on s'intéresse à la segmentation d'images, par rapport aux modélisations par champs de Markov cachés. En effet, dans ce dernier cas de telles simulations ne peuvent être réalisées que par des procédures itératives, de type de l'échantillonneur de Gibbs, coûteuses en temps calcul [21].

## 2.4. PROBABILITÉS FORWARD-BACKWARD

Nous rappelons dans cette sous-section la définition des probabilités Forward-Backward qui jouent un rôle crucial aussi bien au niveau de l'estimation des paramètres qu'à celui de la restauration proprement dite.

**Procédure Forward :**

$\alpha_t(i)$  désigne la probabilité suivante :

$$\alpha_t(i) = P[X_t = \omega_i, Y_1 = y_1, \dots, Y_t = y_t] \quad (12)$$

Cette expression peut se calculer de manière récursive :

– Initialisation :  $\pi_i f_i(y_1)$  pour  $1 \leq i \leq k$

– Induction :  $\alpha_{t+1}(j) = [\sum_{i=1}^k \alpha_t(i) a_{ij}] f_j(y_{t+1})$  pour  $1 \leq j \leq k$ ,  $1 \leq t \leq n - 1$

**Procédure Backward :**

Le calcul est similaire à la démarche précédente. La probabilité backward  $\beta_t(i)$  est définie par :

$$\beta_t(i) = P[Y_{t+1} = y_{t+1}, \dots, Y_n = y_n / X_t = \omega_i] \quad (13)$$

Cette expression donne la probabilité conjointe de la séquence partielle de l'observation de l'instant  $t+1$  jusqu'à l'instant final  $n$ , sachant qu'à l'instant  $t$ , le système est dans la classe  $\omega_i$ . Le calcul de  $\beta_t$  s'obtient à partir de  $\beta_{t+1}$  par une récurrence régressive :

- Initialisation :  $\beta_n(i) = 1$  pour  $1 \leq i \leq k$
- Induction :  $\beta_t(i) = \left[ \sum_{j=1}^k a_{ij} f_j(y_{t+1}) \beta_{t+1}(j) \right]$  pour  $1 \leq i \leq k$ ,  
 $1 \leq t \leq n-1$

Le calcul respectif des probabilités forward et backward se heurte à des difficultés d'ordre numérique. En effet, les quantités figurant dans les expressions de  $\alpha_t(i)$  et  $\beta_t(i)$  sont très petites et la présence des sommes dans ces expressions ne permet pas d'utiliser le logarithme. Afin de s'affranchir de ce problème de dépassement de capacité, Devijver et al. [18] ont proposé une solution qui consiste à remplacer les probabilités conjointes par les probabilités a posteriori. Les reformulations de  $\alpha_t(i)$  et  $\beta_t(i)$  sont les suivantes :

$$\alpha_t^*(i) = P[X_t = \omega_i / Y_1 = y_1, \dots, Y_t = y_t] \quad (14)$$

$$\beta_t^*(i) = \frac{P[Y_{t+1} = y_{t+1}, \dots, Y_n = y_n / X_t = \omega_i]}{P[Y_{t+1} = y_{t+1}, \dots, Y_n = y_n / Y_1 = y_1, \dots, Y_t = y_t]} \quad (15)$$

La procédure de calcul de la probabilité forward devient (l'astérisque sera omis dans ce qui suit) :

- Initialisation :  $\alpha_1(i) = \frac{\pi_i f_i(y_1)}{\sum_{j=1}^k \pi_j f_j(y_1)}$  pour  $1 \leq i \leq k$

- Induction :  $\alpha_{t+1}(i) = \frac{f_i(y_{t+1}) \sum_{j=1}^k \alpha_t(j) a_{ji}}{\sum_{l=1}^k f_l(y_{t+1}) \sum_{j=1}^k \alpha_t(j) a_{jl}}$   
pour  $1 \leq i \leq k, 1 \leq t \leq n-1$ .

La procédure de calcul de la probabilité backward devient :

- Initialisation :  $\beta_n(i) = 1$  pour  $1 \leq i \leq k$

- Induction :  $\beta_t(i) = \frac{\sum_{j=1}^k a_{ij} f_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{l=1}^k f_l(y_{t+1}) \sum_{j=1}^k \alpha_t(j) a_{jl}}$   
pour  $1 \leq i \leq k, 1 \leq t \leq n-1$

### 3. Algorithmes de reconnaissances

Nous rappelons dans cette section deux méthodes Bayésiennes, correspondant à deux fonctions de perte différentes, de la restauration de  $X$ .

#### 3.1. MÉTHODE MODE DE LA MARGINALE A POSTERIORI (MPM)

Considérons la probabilité que le système soit à l'état  $\omega_i$  à l'instant  $t$  sachant l'observation  $Y = y$  (probabilité marginale a posteriori) :

$$\xi_t(i) = P[X_t = \omega_i / Y = y] \quad (16)$$

Cette probabilité peut s'exprimer simplement en fonction des probabilités forward et backward. En effet, nous avons :

$$\xi_t(i) = \frac{P[X_t = \omega_i, Y = y]}{P[Y = y]} = \alpha_t(i) \beta_t(i) \quad (17)$$

La méthode MPM consiste en recherche, pour chaque  $t$ , de l'état  $\omega_j$  qui maximise la probabilité a posteriori  $\xi_t(i)$  :

$$[\hat{X}_t = \omega_j] \Leftrightarrow [\xi_t(j) = \max_{1 \leq i \leq k} \xi_t(i)] \quad (18)$$

L'algorithme MPM se déroule de la manière suivante :

- Calcul en chaque  $t$  et pour chaque  $i$  :
  - la probabilité forward  $\alpha_t(i)$ ,
  - la probabilité backward  $\beta_t(i)$ ,
  - la probabilité a posteriori  $\xi_t(i) = \alpha_t(i) \beta_t(i)$
- Estimation de la réalisation de chaque  $X_t$  par l'état qui maximise  $\xi_t$ .

#### 3.2. MÉTHODE MAXIMUM A POSTERIORI (MAP)

Le principe de la méthode MAP est le suivant :

$$\{\hat{X} = \hat{x}\} \Leftrightarrow \{P[X = \hat{x} / Y = y] = \max_{x \in \Omega^n} P[X = x / Y = y]\} \quad (19)$$

La solution analytique est donnée par l'algorithme de Viterbi [20].

##### Remarque 1

MPM et MAP sont des méthodes globales de recherche de la réalisation de  $X$ , i.e. qui tiennent compte de toute l'information contenue dans la séquence d'observations. Le modèle permet le calcul analytique des solutions, ce qui constitue le principal avantage, en termes de rapidité des calculs, par rapport aux modélisations par champs de Markov cachés. Dans le cadre de ces dernières les solutions de MPM et MAM sont approchées par les algorithmes itératifs de Marroquin et al. [29] et le recuit simulé de Geman et al. [21] respectivement.

### 4. Algorithmes d'estimation des paramètres du modèle

Le problème de l'estimation des paramètres est posé lorsqu'on souhaite rendre les algorithmes de restauration « non supervisés », à savoir applicables quelles que soient les valeurs des paramètres définissant le modèle. Lorsque le nombre de classes est  $k$ , les paramètres définissant la loi de  $X$  sont au nombre de  $k^2$  (ce sont les coefficients  $c_{ij}$  définis par (2), section 2.1). Lorsque les densités  $f_1, f_2, \dots, f_k$  (voir section 2.2) sont gaussiennes, ce que nous supposons dans la suite, les paramètres définissant les lois de  $Y$  conditionnelles à  $X$  sont au nombre de  $2k$  ( $k$  moyennes

et  $k$  variances). Ces paramètres varient, en principe, avec le problème étudié, mais ils peuvent également prendre des valeurs significativement différentes pour un type de problème donné. Ainsi dans l'exemple concernant l'établissement d'une carte (ou segmentation d'une image, section 2) les quatre coefficients  $c_{ij}$  sont déterminés par la taille des zones « eau » et « forêt » et les densités  $f_1, f_2$  (données par moyennes et variances) sont déterminées par le type de végétation, la saison, la nature de l'eau, la météorologie,...

Finalement, en notant  $\theta$  l'ensemble des paramètres déterminant le modèle, le problème préalable à la restauration est celui de l'estimation de  $\theta$  à partir de la seule réalisation  $Y = y$ . Ce problème fait partie du problème général de l'estimation des composantes d'un mélange de distributions, dont les techniques récentes de traitement consistent à utiliser des procédures itératives.

Nous présentons dans cette section deux nouvelles méthodes d'estimation des paramètres dans les chaînes de Markov cachées et les comparons avec la méthode EM. Plus précisément les algorithmes généraux considérés sont :

- L'algorithme Expectation-Maximisation (EM), proposé par Baum et al. [4] connu sous le nom « Forward-Backward Algorithm ».
- L'algorithme Iterative Conditional Estimation (ICE), proposé par Pieczynski [12], [33], [34] dont l'application aux chaînes de Markov a été rapidement décrite dans [5], [19].
- L'algorithme Stochastic Expectation-Maximisation (SEM), proposé par Celeux et al. [14], [30].

#### 4.1. ALGORITHME EM

Le principe de l'algorithme EM classique consiste à maximiser, par rapport aux paramètres du modèle, la fonction de vraisemblance de l'observation  $Y = y$ . C'est une méthode itérative déterministe qui permet de définir, à partir d'une valeur initiale  $\theta^0$  des paramètres, une suite  $(\theta^q)_{q \in \mathbb{N}}$  dont la vraisemblance en  $Y = y$  est une fonction croissante. Il en résulte que la suite  $(\theta^q)_{q \in \mathbb{N}}$  converge vers un paramètre en lequel ladite vraisemblance atteint un maximum local.

Nous décrivons brièvement ci-dessous les itérations de l'algorithme EM dans le cas du modèle des chaînes de Markov cachées [18].

Soit :

$$\Psi_t(i, j) = P[X_t = \omega_i, X_{t+1} = \omega_j / Y = y] \quad (20)$$

la probabilité conjointe d'être à l'instant  $t$  dans la classe  $\omega_i$  et à l'instant suivant dans la classe  $\omega_j$  sachant la suite des observations.

$$\Psi_t(i, j) = \frac{P[X_t = \omega_i, X_{t+1} = \omega_j, Y = y]}{P[Y = y]} \quad (21)$$

L'expression (20), (21) peut s'écrire en fonction des probabilités forward-backward [18] :

$$\Psi_t(i, j) = \frac{\alpha_t(i) a_{ij} f_j(y_{t+1}) \beta_{t+1}(j)}{\sum_{l=1}^k f_t(y_{t+1}) \sum_{j=1}^k \alpha_t(j) a_{jl}} \quad (22)$$

Nous avons défini précédemment la probabilité d'avoir la classe  $\omega_i$  à l'instant  $t$ , sachant l'observation  $Y = y$ , notée  $\xi_t(i)$ . Cette probabilité peut également s'écrire en fonction des  $\Psi_t(i, j)$  :

$$\xi_t(i) = \sum_{j=1}^k \Psi_t(i, j) \quad (23)$$

Les formules de la re-estimation des paramètres du modèle à l'itération  $q + 1$  sont les suivantes [18], [39] :

- Les probabilités initiales et la matrice de transition :

$$\pi_i^{q+1} = \frac{1}{n} \sum_{t=1}^n \xi_t^q(i) \quad a_{ij}^{q+1} = \frac{\sum_{t=1}^{n-1} \Psi_t^q(i, j)}{\sum_{t=1}^{n-1} \xi_t^q(i)} \quad (24)$$

ce qui donne  $c_{ij}^{q+1}$  (voir (3), (4)).

- Le paramètre définissant la densité de probabilité, pour le cas Gaussien considéré, associée à la classe  $\omega_i$ , est donné par le couple  $(\mu_i, \sigma_i^2)$ ,  $\mu_i$  étant la moyenne de la loi et  $\sigma_i^2$  sa variance. On obtient pour la re-estimation :

$$\mu_i^{q+1} = \frac{\sum_{t=1}^n \xi_t^q(i) y_t}{\sum_{t=1}^n \xi_t^q(i)} \quad (\sigma_i^2)^{q+1} = \frac{\sum_{t=1}^n \xi_t^q(i) (y_t - \mu_i^{q+1})^2}{\sum_{t=1}^n \xi_t^q(i)} \quad (25)$$

L'algorithme EM se déroule de la façon suivante :

- Initialisation des paramètres  $a_{ij}^0, \mu_i^0, (\sigma_i^2)^0$  pour  $1 \leq i, j \leq k$
- A chaque itération  $q$  :
  - Etape « E » : calcul des probabilités suivantes :
    - $\alpha_t(i)$  et  $\beta_t(i)$
    - Déduction de  $\Psi_t^q(i, j)$  à partir de  $\alpha_t(i)$  et  $\beta_t(i)$  calculée sur la base des paramètres  $a_{ij}^q, \mu_i^q, (\sigma_i^2)^q$
  - Etape « M » : calcul des paramètres  $\pi_i^{q+1}, a_{ij}^{q+1}, \mu_i^{q+1}, (\sigma_i^2)^{q+1}$  par les formules ci-dessus.

#### Remarque 2

Le choix de l'initialisation des paramètres du modèle peut avoir une forte influence sur la rapidité de la convergence de l'algorithme EM. Par ailleurs il existe un risque de convergence vers un maximum local éloigné du maximum absolu de la fonction de vraisemblance de l'observation  $y$ .

## 4.2. ALGORITHME ICE

### 4.2.1. Principe général

L'algorithme ICE est une procédure récente générale d'estimation des paramètres dans le cas des données cachées [33], [34]. Son principe est de considérer un estimateur  $\hat{\theta} = \hat{\theta}(X, Y)$  défini à partir des données complètes  $(X, Y)$  et de l'approcher par une fonction de  $Y$ , les seules données observables. La meilleure approximation, au sens de l'erreur quadratique moyenne, est l'espérance conditionnelle. L'espérance conditionnelle de  $\hat{\theta} = \hat{\theta}(X, Y)$  par rapport à  $Y$ , notée  $E[\hat{\theta}(X, Y)/Y]$ , dépend du paramètre  $\theta$ . Pour la calculer, on utilise une démarche itérative qui consiste à calculer la valeur suivante des paramètres en fonction de leur valeur « courante », i.e. :

$$\theta^{q+1} = E_{\theta^q}[\hat{\theta}(X, Y)/Y = y] \quad (26)$$

où  $E_{\theta^q}$  est l'espérance conditionnelle correspondant à  $\theta^q$ .

Ainsi la procédure de l'algorithme ICE est la suivante :

- L'initialisation du paramètre  $\theta$ ,
- $\theta^{q+1}$  est calculé à partir de  $\theta^q$  et  $Y = y$   
par  $\theta^{q+1} = E_{\theta^q}[\hat{\theta}(X, Y)/Y = y]$

Lorsque  $E_{\theta^q}[\hat{\theta}/Y = y]$  n'est pas explicitement calculable mais la simulation des réalisations de  $X$  selon la loi conditionnelle à  $Y$  possible, on peut utiliser une approximation stochastique. En effet, l'espérance conditionnelle est aussi l'espérance selon la loi conditionnelle et peut être approchée, en vertu de la loi des grands nombres, par la moyenne empirique. On peut ainsi simuler  $N$  réalisations  $x_1, \dots, x_N$  de  $X$  selon la loi conditionnelle à  $Y = y$  donnée par  $\theta^q$  et poser :

$$\theta^{q+1} = \frac{1}{N} [\hat{\theta}(x_1, y) + \dots + \hat{\theta}(x_N, y)] \quad (27)$$

### 4.2.2. Cas des chaînes de Markov cachées

La procédure ICE est applicable au problème de l'estimation des paramètres dans les chaînes de Markov cachées. Nous examinons successivement le problème de l'estimation des paramètres définissant la loi de  $X$  et celui de l'estimation des paramètres du bruit.

Re-estimation des paramètres de la loi  $X$  :

Rappelons la définition des  $c_{ij}$  :

$$c_{ij} = P[X_m = \omega_i, X_{m+1} = \omega_j] \quad (28)$$

Supposons, conformément au principe général de ICE, que  $X$  est observable. Le paramètre de transition  $c_{ij}$  peut alors être estimé par :

$$\hat{c}_{ij}(X) = \frac{1}{n-1} \sum_{t=1}^{n-1} 1_{[X_t = \omega_i, X_{t+1} = \omega_j]} \quad (29)$$

qui est la fréquence empirique. Afin d'obtenir la valeur suivante du paramètre  $c_{ij}$  on considère, conformément au principe général de ICE, l'espérance conditionnelle de  $\hat{c}_{ij}$  :

$$c_{ij}^{q+1} = E_{\theta^q}[\hat{c}_{ij}(X)/Y = y] \quad (30)$$

Ce qui donne ( $P^q$  est définie par  $\theta^q$ ) :

$$c_{ij}^{q+1} = \frac{1}{n-1} \sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j/Y = y] \quad (31)$$

La matrice de transition correspondante à l'itération  $q+1$  est donc

$$a_{ij}^{q+1} = \frac{c_{ij}^{q+1}}{\sum_{j=1}^k c_{ij}^{q+1}} = \frac{\sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j/Y = y]}{\sum_{j=1}^k \sum_{t=1}^{n-1} P^q[X_t = \omega_i, X_{t+1} = \omega_j/Y = y]} \quad (32)$$

ou encore, en reprenant les notations ci-dessus :

$$a_{ij}^{q+1} = \frac{\sum_{t=1}^{n-1} \Psi_t^q(i, j)}{\sum_{t=1}^{n-1} \xi_t^q(i)} \quad (33)$$

### Remarque 3

Les formules ci-dessus de la re-estimation des paramètres correspondant à la loi de  $X$  sont les mêmes que celles obtenues par la méthode EM. Contrairement à EM, la méthode ICE n'utilise pas la notion de vraisemblance et le calcul reste valable quelle que soit la forme du bruit. Notons cependant l'absence des résultats théoriques concernant le comportement de la suite aléatoire définie par le principe de ICE.

Re-estimation des paramètres des lois de  $Y_t$  conditionnelles à  $X_t$  : Dans le cas gaussien considéré les paramètres des densités conditionnelles sont les moyennes  $\mu_i$  et les variances  $\sigma_i^2$ . Suivant le principe de ICE supposons  $X$  observable. Les paramètres  $\mu_i$  et  $\sigma_i^2$  peuvent alors être estimés par les moyennes et les variances empiriques :

$$\hat{\mu}_i(X, Y) = \frac{\sum_{j=1}^n Y_j 1_{[X_j = \omega_i]}}{\sum_{j=1}^n 1_{[X_j = \omega_i]}} \quad \hat{\sigma}_i^2(X, Y) = \frac{\sum_{j=1}^n (Y_j - \hat{\mu}_i)^2 1_{[X_j = \omega_i]}}{\sum_{j=1}^n 1_{[X_j = \omega_i]}} \quad (34)$$

Les espérances conditionnelles des estimateurs  $\hat{\mu}_i(X, Y)$ ,  $\hat{\sigma}_i^2(X, Y)$  ne sont pas calculables explicitement. On peut cependant utiliser l'approximation stochastique : en effet, nous avons vu dans la section 2.3 que la simulation des réalisations de  $X$ , selon sa loi a posteriori, est possible.

Finalement la démarche à chaque itération  $q$  est la suivante :

- Simulation de  $N$  réalisation de  $X$  selon sa loi a posteriori :  $x^1, \dots, x^N$  (chaque  $x^j$  est un  $n$ -uple)
- Calcul, pour chaque  $1 \leq i \leq k$  et  $1 \leq j \leq N$ , de  $\mu_i^j$  et  $(\sigma_i^j)^2$  (à partir de  $x^j, y$ ) par les formules (34).

– Calcul, pour chaque  $1 \leq i \leq k$ , de  $\mu_i^{q+1}$  et  $(\sigma_i^{q+1})^2$  par :

$$\begin{aligned} \mu_i^{q+1} &= \frac{1}{N}(\mu_i^1 + \dots + \mu_i^N) \\ (\sigma_i^{q+1})^2 &= \frac{1}{N}((\sigma_i^1)^2 + \dots + (\sigma_i^N)^2) \end{aligned} \quad (35)$$

#### 4.2.3. ICE modifié (MICE)

Nous proposons une variante originale de l'ICE suivante : au lieu de faire des simulations des réalisations de  $X$  selon sa loi a posteriori on effectue, pour chaque instant  $t$ , un tirage dans  $\Omega$  selon la loi a posteriori marginale donnée par  $\xi_t^q(i) = P[X_t = \omega_i / Y = y]$ . Le reste de la démarche demeure identique.

#### Remarque 4

ICE est une méthode d'estimation à la fois déterministe (les paramètres  $a_{ij}$ ) et stochastique (les paramètres  $\mu_i$  et  $\sigma_i^2$ ). Grâce à sa phase stochastique, ICE s'avère moins sensible au choix des paramètres de l'initialisation que l'algorithme EM.

### 4.3. ALGORITHME SEM

#### 4.3.1. Principe du SEM

L'algorithme SEM est une variante de l'algorithme EM. C'est une procédure itérative permettant de définir une suite  $(\theta^q)_{q \in N}$  à partir des observations  $Y = y$  et des réalisations de  $X$ , obtenues par des tirages stochastiques. La procédure se déroule de la façon suivante :

- Initialisation des paramètres à estimer.
- A chaque itération  $q$ ,
  - Simulation d'une seule réalisation  $x$  de  $X$  selon la loi a posteriori basée sur les paramètres courants (même démarche que dans le cas de ICE pour l'estimation des paramètres du bruit, avec  $N = 1$ )
  - Re-estimation des paramètres  $c_{ij}$  par :
 
$$c_{ij}^{q+1} = \frac{1}{n-1} \sum_{t=1}^{n-1} 1_{[X_t = \omega_i, X_{t+1} = \omega_j]}$$
 (tout se passe comme si le résultat du tirage était la vraie réalisation de  $X$ ).
  - Re-estimation des paramètres du bruit par la même démarche que celle de ICE.

#### 4.3.2. SEM modifié (MSEM)

Nous proposons la modification originale suivante de l'algorithme SEM : la démarche générale est la même sauf en ce qui concerne la manière de simuler, à chaque itération, la réalisation de  $X$ . Nous avons vu que dans la version originale, cette réalisation était obtenue par le tirage selon la loi de  $X$  a posteriori. Dans la version modifiée, on considère les composantes du vecteur aléatoire  $X$  deux par deux :  $(X_1, X_2), (X_3, X_4), \dots, (X_{n-1}, X_n)$  (nous

supposons  $n$  pair). La réalisation de chaque couple  $(X_i, X_{i+1})$  est alors simulée selon sa loi conditionnelle à  $Y = y$  (donnée par (21), (22)) et de façon indépendante. Les réalisations de tous les couples donnent la réalisation de  $X$ .

Cette version s'avère légèrement plus performante que la version originale. Elle est cependant plus longue. En effet, le calcul des lois des couples ci-dessus conditionnelles à  $Y = y$  fait appel à la procédure « backward-forward », ce qui n'est pas le cas lorsqu'on utilise la version originale qui utilise la seule procédure « backward ».

#### Remarque 5

Les études numériques montrent une légère supériorité de MICE et MSEM sur ICE et SEM respectivement dans le contexte des chaînes de Markov cachées [6]. Cette situation pourrait paraître paradoxale dans la mesure où les démarches de MICE et MSEM s'écartent de la théorie générale. Notons cependant que ce phénomène pourrait cacher des propriétés profondes propres à toutes les méthodes d'estimation utilisant des observations obtenues « artificiellement » par tirages sur la base de paramètres courants. En effet, ces observations sont « plus indépendantes » dans les cas de MICE et MSEM que dans ceux de ICE et SEM. Elles contiennent donc, intuitivement, plus d'information.

#### Remarque 6

Lorsqu'on limite le nombre de tirages (à chaque itération) intervenant dans ICE à un seul, ICE et SEM sont identiques pour l'estimation des paramètres de bruit, et cela s'effectue d'une manière stochastique. Par ailleurs, ICE et EM re-estiment les paramètres de la loi  $X$  de la même façon déterministe. Cette situation est précisée dans le Tableau 1 ci-dessous.

Tableau 1

Paramètres à estimer	EM	ICE	SEM
$(c_{ij})$	Déterministe		Stochastique
$(\mu_i, \sigma_i^2)$	Déterministe	Stochastique	

## 5. Étude de l'efficacité des algorithmes EM, MICE, MSEM

Nous avons testé, dans un premier temps, le comportement des algorithmes EM, ICE et SEM dans différentes situations d'homogénéité et de bruitage des chaînes. Les versions originales des algorithmes ICE et SEM s'avèrent légèrement moins compétitives vis-à-vis de l'algorithme EM dans ce contexte. Différents résultats numériques peuvent être consultés dans [6]. Nous présentons ci-dessous uniquement les résultats correspondant à MICE et MSEM dont l'efficacité est comparable, et parfois supérieure, à celle de EM.

Notons que cette situation ne préjuge pas de la supériorité de EM sur ICE et SEM en général. Nous montrons dans [31] que dans le contexte de la segmentation bayésienne non supervisée locale (chaque pixel est classé à partir de l'observation sur un contexte de petite taille) les performances de ces trois méthodes sont comparables. ICE s'avère également compétitive vis-à-vis de EM dans le contexte de champs de Markov cachés [34].

Nous testons les performances des algorithmes d'estimations en présence à partir des simulations des différentes chaînes de Markov bruitées. Nous avons considéré plusieurs cas selon la forme du bruit, l'homogénéité de la chaîne et le niveau de bruitage. Rappelons que la distribution conditionnelle de chaque  $Y_t$  sachant  $X_t = \omega_i$  est donnée par une densité gaussienne de moyenne  $\mu_i$  et de variance  $(\sigma_i)^2$ . Pour chaque  $t$ , nous avons :

$$X_t = \omega_i \Rightarrow Y_t \text{ suit la loi } N(\mu_i, \sigma_i^2)$$

On peut distinguer trois types de bruits selon leurs moyennes et variances respectives :

1.  $\mu_1 \neq \mu_2, \sigma_1 = \sigma_2$  : bruit « MD » (discrimination par les moyennes)
2.  $\mu_1 = \mu_2, \sigma_1 \neq \sigma_2$  : bruit « VD » (discrimination par les variances)
3.  $\mu_1 \neq \mu_2, \sigma_1 \neq \sigma_2$  : bruit « MVD » (discrimination par les variances et les moyennes)

Dans les applications pratiques les bruitages sont généralement de type MVD, les types MD et VD en étant les cas « extrêmes ». Afin de tester l'influence du type de bruit sur le comportement des divers algorithmes, nous considérerons ces cas « extrêmes ».

L'homogénéité de la chaîne est le deuxième facteur dont nous étudions l'influence sur le comportement des méthodes en présence. Nous considérons deux cas :

1. Chaîne homogène :  $c_{22} = c_{11} = 0.49, c_{21} = c_{12} = 0.01$ .

La matrice de transition est alors donnée par :

$$a_{22} = a_{11} = 0.98, a_{21} = a_{12} = 0.02.$$

Dans la suite ce cas sera dit « H » et la chaîne sera appelée chaîne A.

2. Chaîne non homogène :  $c_{22} = c_{11} = 0.4, c_{21} = c_{12} = 0.1$ .

La matrice de transition est alors donnée par :

$$a_{22} = a_{11} = 0.8, a_{21} = a_{12} = 0.2.$$

Dans la suite ce cas sera dit « NH » et la chaîne sera appelée chaîne B.

Nous aurons ainsi quatre familles de cas : MDH, MVH, MDNH, MVNH. Dans chacune de ces familles nous étudions différents niveaux de bruitage. Dans le cas MD, ce niveau sera défini par l'écart entre les moyennes  $\delta = m_2 - m_1$  (nous supposons  $m_1 < m_2$ ). Dans le cas VD, il le sera par  $\rho = \frac{\sigma_2^2}{\sigma_1^2}$ .

Enfin,  $n$  désigne la taille de l'échantillon.

Afin d'étudier les performances des estimateurs, nous avons comparé les valeurs estimées à partir des données bruitées avec

celles calculées à partir des données complètes. Ces dernières seront appelées « valeurs empiriques » et notées « VE ». Le calcul des valeurs empiriques se fait par les formules suivantes :

– Les paramètres  $c_{ij}$

$$c_{ij}^* = \frac{\sum_{t=1}^{n-1} 1_{[X_t=\omega_i, X_{t+1}=\omega_j]}}{n-1} \quad (36)$$

– La moyenne correspondant à la classe  $\omega_i$  :

$$\mu_i^* = \frac{\sum_{t=1}^n y_t 1_{[X_t=\omega_i]}}{\text{Card}(Q_i)} \quad (37)$$

– La variance correspondant à la classe  $\omega_i$  :

$$(\sigma_i^*)^2 = \frac{\sum_{t=1}^n (y_t - \mu_i^*)^2 1_{[X_t=\omega_i]}}{\text{Card}(Q_i)} \quad (38)$$

où  $Q_i$  est l'ensemble des  $t$  tels que  $X_t = \omega_i$ .

#### Remarque 7

Les valeurs empiriques VE diffèrent des vraies valeurs (valeurs de simulation), notées VV, dans le cas des petits échantillons. Lorsque l'échantillon est suffisamment riche, VV sont proches des VE (loi des grands nombres). Nos résultats correspondent aux échantillons de taille 1000.

## 5.1. INITIALISATION DES ALGORITHMES

Pour  $k$  nombre de classes, les paramètres sont initialisés de la façon suivante :

Initialisation des paramètres de la loi a priori :

$$\begin{aligned} & - \forall 1 \leq i \leq k \quad \pi_i^0 = \frac{1}{k} \\ & - \forall 1 \leq i, j \leq k \text{ avec } i \neq j \quad a_{ij}^0 = \frac{1}{2(k-1)} \\ & - \forall 1 \leq i \leq k \quad a_{ii}^0 = 0.5 \end{aligned}$$

Initialisation des moyennes du bruit :

Pour  $k$  pair et  $\forall i$  tel que  $0 \leq i \leq \frac{k}{2}$  :

$$\begin{aligned} & - \mu_{i+1}^0 = M - \left(\frac{k}{2} - i\right) \frac{\Sigma}{2} \\ & - \mu_{k-i}^0 = M + \left(\frac{k}{2} - i\right) \frac{\Sigma}{2} \end{aligned}$$

Pour  $k$  impair et  $\forall i$  tel que  $0 \leq i < k$  :

$$- \mu_{i+1}^0 = M - \left(\frac{k}{2} - i\right) \frac{\Sigma}{2}$$

Initialisation des variances du bruit :

$$- \forall 1 \leq i \leq k \quad \sigma_i = \Sigma$$

sachant que  $M$  représente la moyenne et  $\Sigma$  l'écart-type de l'échantillon.

## 5.2. LE CRITÈRE D'ARRÊT

Le critère d'arrêt des procédures d'estimation peut être fonction de l'objectif final d'utilisation des paramètres estimés. Dans le cas étudié ici, cet objectif est la segmentation d'images par la méthode MPM. La qualité finale des résultats obtenus est fonction de deux phénomènes indépendants : la qualité de l'estimation et la robustesse de la méthode MPM par rapport aux paramètres. Lorsque MPM est robuste par rapport à un paramètre dans le sens où sa qualité baisse peu lorsque la valeur estimée de ce paramètre s'écarte de sa vraie valeur, on peut se contenter d'une valeur approximative dudit paramètre. A contrario, si MPM est peu robuste par rapport à un paramètre, on doit être exigeant vis-à-vis de la qualité de son estimation. Nous avons étudié les courbes de robustesse de la méthode MPM [6] et il s'avère que les moyennes et les variances du bruit sont des paramètres importants. Nous proposons ci-dessous un critère d'arrêt basé sur l'évolution de ces paramètres.

Soit  $\lambda^q$  le paramètre estimé à l'itération  $q$  et  $\lambda^{q,m}$  la moyenne sur les  $m$  dernières itérations. Notre critère d'arrêt est basé sur les fluctuations des paramètres au cours des itérations. Soit la distance  $D$  définie par :

$$D^2(\lambda^q, \lambda^{q,m}) = \frac{1}{k} \sum_{i=1}^k c_i (\lambda_i^q - \lambda_i^{q,m})^s \quad (39)$$

avec  $s = 2$  si  $\lambda^q = \mu^q$  (estimation des moyennes) et  $s = 1$  si  $\lambda^q = \sigma^q$  (estimation des variances). Les coefficients de pondération  $c_i$  sont liés à la robustesse de la méthode de segmentation MPM. Cette méthode étant plus sensible aux moyennes qu'aux variances, nous avons choisi les valeurs des coefficients  $c_i$  correspondants, égales à 0.9 et 0.1 respectivement. La procédure de l'estimation s'arrête lorsque la distance  $D$  est inférieure à un seuil  $\varepsilon$  strictement positif pré défini. Nous avons déterminé, de façon expérimentale, que  $\varepsilon = 0.01$  convenait pour la chaîne A et que  $\varepsilon = 0.001$  était nécessaire pour la chaîne B. Dans le cas des images (section 7) on peut se contenter de  $\varepsilon = 0.1$  (des seuils plus petits n'apportent pas d'amélioration des résultats des segmentations).

## 5.3. RÉSULTATS DES EXPÉRIMENTATIONS

Nous considérons un échantillon de taille 1000 et le nombre de classes égal à 2 (les algorithmes restent valables pour un nombre quelconque de classes). Les algorithmes MICE, EM et MSEM sont testés sur les chaînes A et B, chacune étant bruitée par un bruit MD et un bruit VD. Les 3 algorithmes sont initialisés par la même méthode d'initialisation décrite ci-dessus. Nous présentons uniquement les résultats correspondant aux bruitages relativement forts, ceux correspondant aux rapports signal sur bruit plus importants pouvant être consultés dans [6]. Les valeurs estimées sont celles obtenues à l'arrêt automatique des algorithmes.

Les résultats des estimations sont présentés dans les Tableaux 2-5. Bien que la restauration ne fasse pas l'objet de cette sous-section nous donnons dès maintenant, dans la dernière ligne des tableaux, le taux des pixels mal classés par la méthode MPM utilisant les valeurs estimées des paramètres.

### Notations

VE : Les valeurs empiriques.

MSEM : Les valeurs estimées par MSEM.

EM : Les valeurs estimées par EM.

MICE : Les valeurs estimées par MICE.

$T$  : Nombre d'itérations.

$\tau(\%)$  : Taux des pixels mal classés par la méthode MPM utilisant les VE ou les valeurs estimées des paramètres.

Tableau 2. —  $\mu_1 = 1, \mu_2 = 2$  et  $\sigma_1 = \sigma_2 = 1$  (chaîne A)

	VE	EM	MSEM	MICE
$\mu_1$	0.976	0.976	0.979	0.979
$\mu_2$	2.017	2.013	2.003	2.009
$\sigma_1^2$	1.036	1.031	1.032	1.033
$\sigma_2^2$	0.937	0.944	0.956	0.949
$\pi_1$	0.422	0.420	0.416	0.422
$a_{11}$	0.976	0.978	0.990	0.977
$a_{22}$	0.983	0.984	0.986	0.984
$T$	/	23	32	25
$\tau(\%)$	4.7	4.4	4.4	4.4

Nous constatons, selon les tableaux 2-5, que le comportement des trois algorithmes est excellent, en ce qui concerne le taux de bonne restauration, et qu'il est impossible de les départager. Notons que les bruitages sont importants : ainsi que nous le verrons dans la suite, des images binaires bruitées par ce type de bruit sont pratiquement impossibles à reconnaître visuellement.

### Évolution des valeurs estimées au cours des itérations

Nous présentons sur les Figures 1-8 quelques résultats concernant l'évolution des valeurs estimées par différentes méthodes au cours des itérations. Concernant la chaîne A, nous traitons le cas de deux bruitages de type MD de niveaux différents et concernant la chaîne B, deux bruitages de types MD et VD respectivement. Les résultats complets de ce type peuvent être consultés dans [6].

**Tableau 3.** —  $\mu_1 = 1, \mu_2 = 1$  et  $\sigma_1 = \sigma_2^2 = 3$  (chaîne A)

	VE	EM	MSEM	MICE
$\mu_1$	0.976	0.956	0.968	0.923
$\mu_2$	1.029	1.47	1.041	1.066
$\sigma_1^2$	1.036	1.095	1.129	1.060
$\sigma_2^2$	2.810	2.833	2.890	2.767
$\pi_1$	0.422	0.445	0.471	0.434
$a_{11}$	0.976	0.974	0.979	0.973
$a_{22}$	0.983	0.979	0.985	0.979
$T$	/	59	54	50
$\tau(\%)$	10.8	11.4	11.3	11.6

**Tableau 5.** —  $\mu_1 = \mu_2 = 1$ , et  $\sigma_1 = 1, \sigma_2^2 = 3$  (chaîne B)

	VE	EM	MSEM	MICE
$\mu_1$	0.989	0.946	0.989	0.880
$\mu_2$	1.016	1.055	1.016	1.102
$\sigma_1^2$	1.024	1.100	1.119	0.978
$\sigma_2^2$	2.811	2.731	2.748	2.716
$\pi_1$	0.481	0.488	0.489	0.442
$a_{11}$	0.786	0.912	0.926	0.905
$a_{22}$	0.801	0.919	0.933	0.926
$T$	/	275	400	319
$\tau(\%)$	30.1	30.2	30.2	30.7

**Tableau 4.** —  $\mu_1 = 1, \mu_2 = 2$  et  $\sigma_1 = \sigma_2 = 1$  (chaîne B)

	VE	EM	MSEM	MICE
$\mu_1$	0.976	0.884	0.851	0.844
$\mu_2$	2.017	2.025	2.032	1.942
$\sigma_1^2$	1.036	0.908	0.832	0.904
$\sigma_2^2$	0.937	0.926	0.956	0.984
$\pi_1$	0.481	0.444	0.435	0.399
$a_{11}$	0.786	0.752	0.726	0.761
$a_{22}$	0.801	0.802	0.775	0.841
$T$	/	59	40	71
$\tau(\%)$	24.4	24.8	25.6	24.7

Dans le cas d'une chaîne homogène (chaîne A), nous pouvons observer une certaine similitude des comportements des algorithmes. Leur convergence est par ailleurs rapide, ce qui se traduit par un nombre relativement peu élevé d'itérations. Le cas non homogène (chaîne B) pose davantage de problèmes et met en lumière les différences dans les comportements des algorithmes. Le caractère stochastique de MICE et MSEM apparaît et la con-

vergence des trois méthodes est plus lente. Ainsi l'homogénéité de la chaîne influe sur le comportement des algorithmes d'estimation. Par ailleurs, le type de bruit (MD ou VD) joue également un rôle important : bien que les niveaux de bruitages MD (Figures 5, 6) et VD (Figures 7, 8) soient comparables, le temps de stabilisation des algorithmes est sensiblement plus long dans le cas VD. Notons que cela pourrait être partiellement dû à l'initialisation qui est de type MD (section 5.1). Des courbes analogues donnant l'évolution, dans différentes situations, des valeurs estimées des variances peuvent être consultées dans [6].

En conclusion, le comportement des algorithmes est excellent dans les cas de type (chaîne homogène + bruit MD), les cas de type (chaîne non-homogène + bruit VD) étant plus difficiles à traiter.

## 6. Restauration non supervisée dans le cas des chaînes simulées

Nous étudions dans cette section le comportement de trois algorithmes de restauration non supervisée de chaînes obtenus par l'adjonction à la méthode MPM des algorithmes EM, MICE et MSEM respectivement. L'objectif est de mettre en évidence les différences de leur comportement (vitesse de convergence, stabilité) en fonction de l'homogénéité de la chaîne, rapport signal sur bruit, forme de bruit.

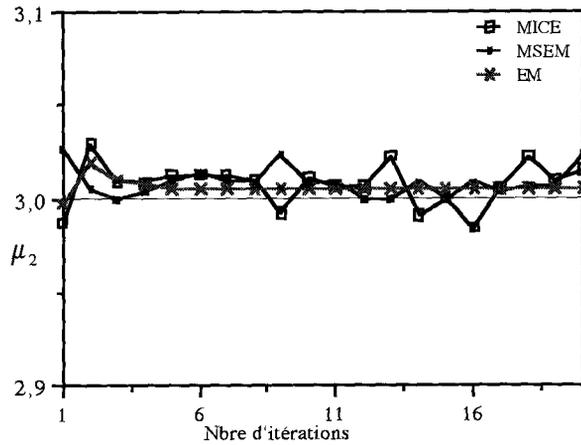
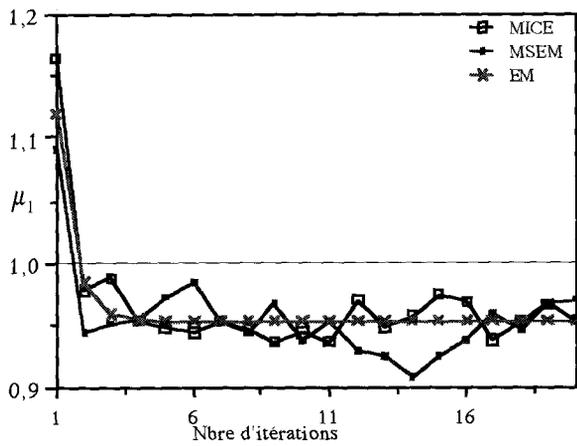


Figure 1, 2 (chaîne A). – Évolution des moyennes  $\mu_1$  et  $\mu_2$  au cours des itérations. Bruit MD :  $\mu_1 = 1, \mu_2 = 3, \sigma_1 = \sigma_2 = 1$

Un comportement similaire est observé pour les variances [6].

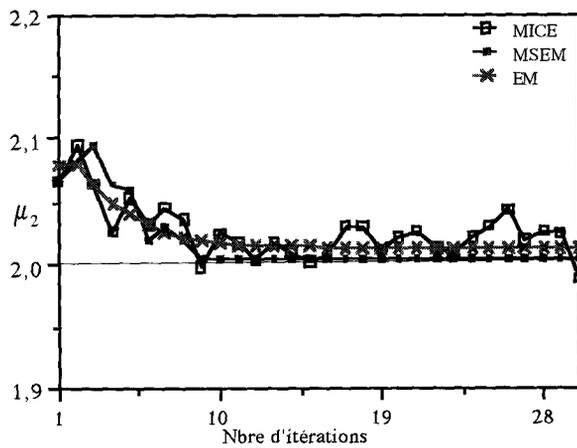
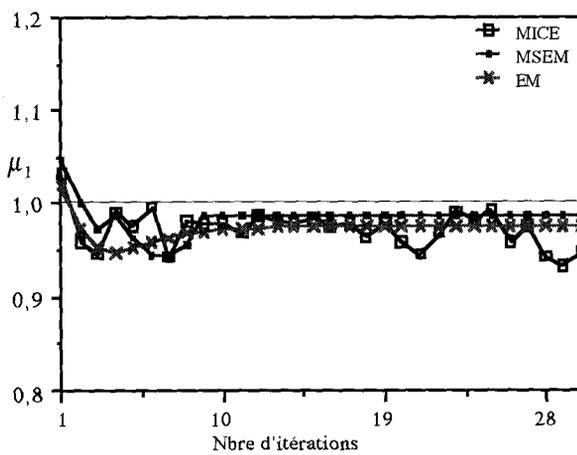


Figure 3, 4 (chaîne A). – Évolution des moyennes  $\mu_1$  et  $\mu_2$  au cours des itérations. Bruit MD :  $\mu_1 = 1, \mu_2 = 2, \sigma_1 = \sigma_2 = 1$

Un comportement similaire est observé pour les variances [6].

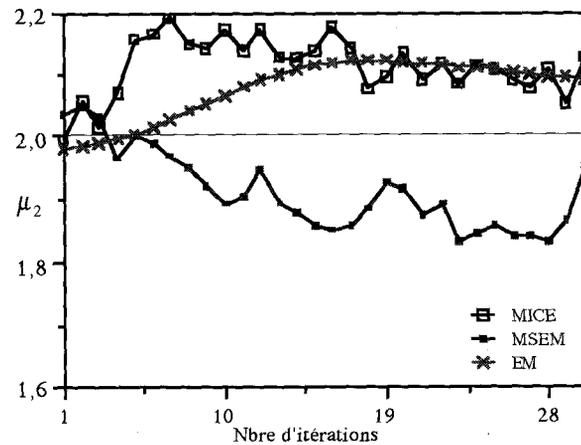
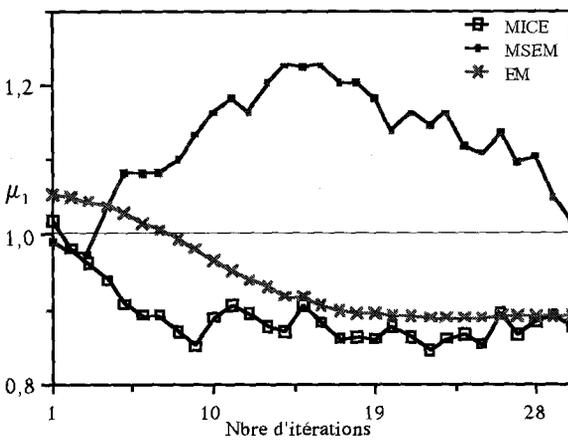


Figure 5, 6 (chaîne B). – Évolution des moyennes  $\mu_1$  et  $\mu_2$  au cours des itérations. Bruit MD :  $\mu_1 = 1, \mu_2 = 2, \sigma_1 = \sigma_2 = 1$

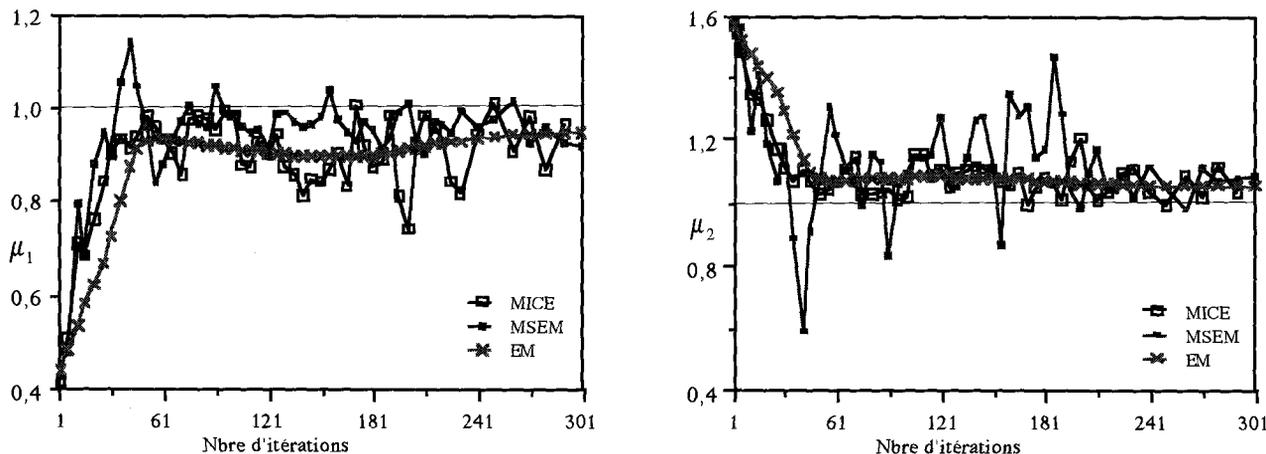


Figure 7, 8 (chaîne B). - Évolution des moyennes  $\mu_1$  et  $\mu_2$  au cours des itérations. Bruit VD :  $\mu_1 = \mu_2 = 1, \sigma_2 = 1, \sigma_1^2 = 3$

### 6.1. ÉVOLUTION DES RÉSULTATS AU COURS DES ITÉRATIONS

Nous considérons quatre bruitages. Trois d'entre eux sont de type MD, avec  $\delta = \Delta m = 3, 2, 1$  respectivement. Le quatrième est de type VD avec  $\rho = \frac{\sigma_2^2}{\sigma_1^2} = 3$ .

Afin d'étudier les performances des différentes méthodes, nous précisons dans le Tableau 6 le taux d'erreur « théorique », i.e. le taux obtenu par le MPM utilisant les valeurs empiriques VE de la chaîne simulée. Ce taux d'erreur représente la borne théorique des taux obtenus par les méthodes non supervisées.

Tableau 6. — Les taux théoriques  $\tau_0$  pour les chaînes A et B

	$\delta = \Delta m = 3$	$\delta = \Delta m = 2$	$\delta = \Delta m = 1$	$\rho = \frac{\sigma_2^2}{\sigma_1^2} = 3$
A	0.4%	1.8%	4.7%	10.8%
B	4.2%	10.7%	24.4%	30.1%

Les Figures 9–16 représentent les courbes décrivant l'évolution du taux de bonne classification, en fonction du nombre d'itérations, donné par les algorithmes EM + HPM, MICE + MPM, MSEM + MPM respectivement.

Ces résultats nous conduisent aux conclusions suivantes :

1. La décroissance de l'homogénéité de la chaîne implique la détérioration des performances de toutes les méthodes en présence. Notons que cette détérioration n'est pas due à la phase « estimation des paramètres » mais à la phase « segmentation » (voir Remarque 8).
2. Lorsque le bruitage de type MD est peu important ( $\delta = \Delta m = 2, 3$ ) le comportement de toutes les méthodes est

satisfaisant. En particulier la convergence est rapide (une dizaine d'itérations).

3. La convergence des méthodes EM+MPM, MICE+MPM et MSEM+MPM est beaucoup plus lente dans le cas VD (environ deux cent cinquante itérations) que dans le cas MD (une trentaine d'itérations).
4. L'algorithme MSEM semble être mis en difficulté, au niveau de la vitesse de convergence, dans le cas non homogène fortement bruité (bruit MD avec  $\delta = 1$  ou VD avec  $\rho = 3$ ). Le choix de EM ou MICE est préférable dans ce contexte.

Nous avons effectué d'autres simulations de ce type [6] dont les résultats corroborent les conclusions énoncées ci-dessus.

#### Remarque 8

Lorsque la chaîne est peu homogène (chaîne B), le problème de la pertinence même d'utilisation de la modélisation par chaînes de Markov cachées se pose. Dans le cas MD avec  $\delta = 3$ , le taux théorique de pixels mal classés par MPM est de 4.2% (Tableau 6). Dans le cas de la classification aveugle (chaque  $x_t$  est estimé à partir du seul  $y_t$ ) ce taux est de 7%. Le même problème se pose dans le cas VD avec  $\rho = 3$  où ces taux sont respectivement de 30.1% (Tableau 6) et 31%. L'utilisation de la classification aveugle, beaucoup plus simple, pourrait s'avérer plus pertinente dans ces cas. Notons que les taux théoriques de la classification aveugle sont les mêmes dans le cas de la chaîne A (l'homogénéité n'intervient pas). Ainsi les taux de la méthode MPM figurant dans le même tableau apportent une preuve manifeste de l'intérêt de la modélisation par chaînes de Markov cachées dans ce contexte. Finalement, l'homogénéité de la chaîne est un facteur important et il pourrait s'avérer pertinent, dans certaines situations, de renoncer à l'utilisation des méthodes non supervisées de type MPM au bénéfice des méthodes plus simples. Ce même phénomène a été observé dans le cadre de segmentations d'images par des méthodes utilisant les modélisations par champs de Markov cachés : dans certains cas les méthodes contextuelles, beaucoup plus simples, s'avèrent plus efficaces [12], [28].

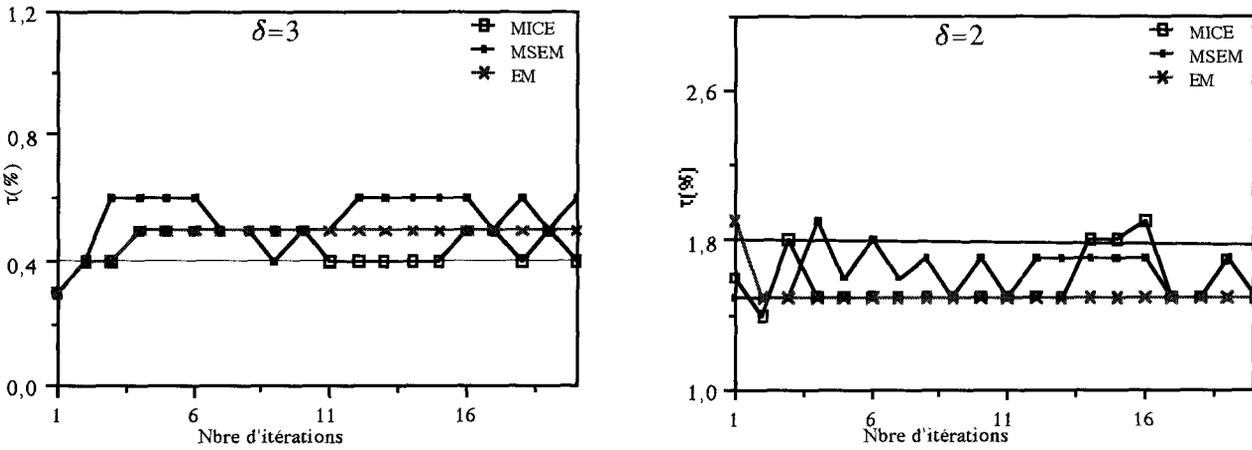


Figure 9, 10 (chaîne A). – Taux d'erreur de MPM en fonction du nombre d'itération des estimateurs EM, MICE, MSEM

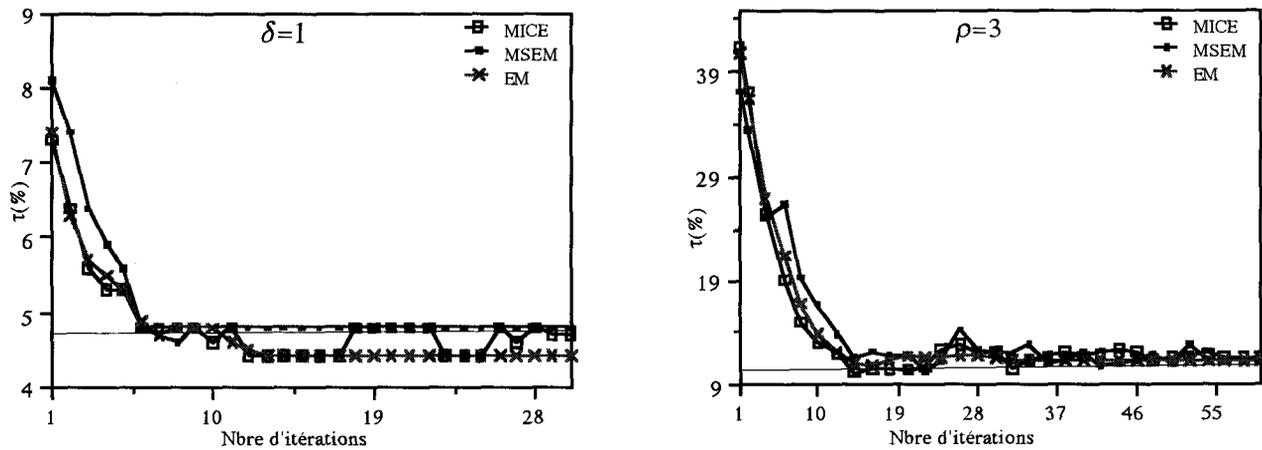


Figure 11, 12 (chaîne A). – Taux d'erreur de MPM en fonction du nombre d'itération des estimateurs EM, MICE, MSEM

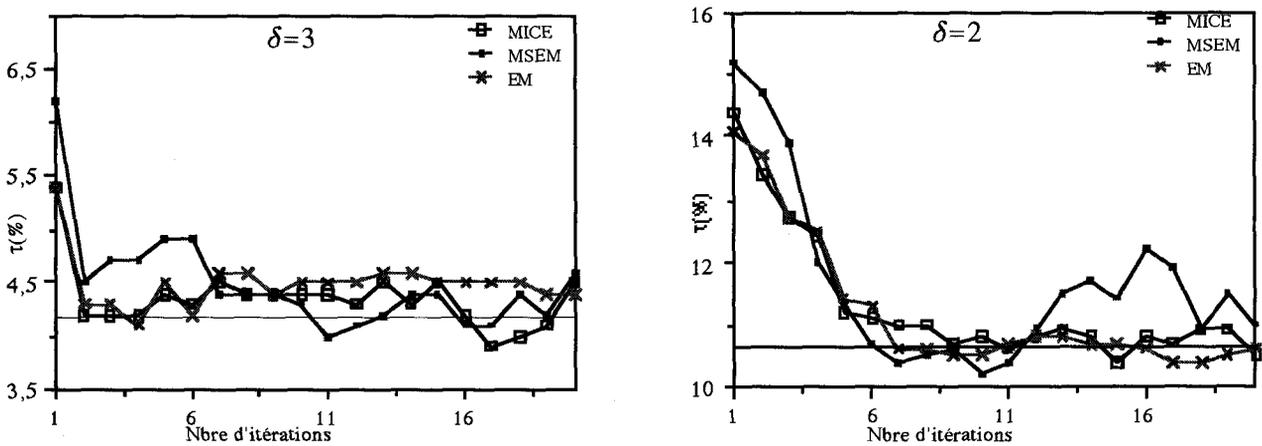


Figure 13, 14 (chaîne B). – Taux d'erreur de MPM en fonction du nombre d'itérations des estimateurs EM, MICE, MSEM

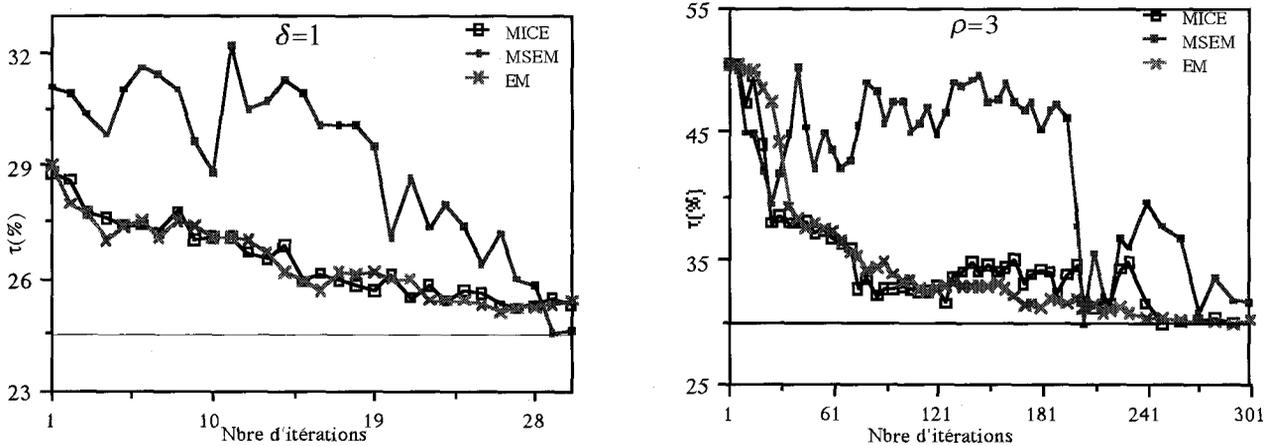


Figure 15,16 (chaîne B). – Taux d'erreur de MPM en fonction du nombre d'itérations des estimateurs EM, MICE, MSEM

## 6.2. AUGMENTATION DU BRUIT

Nous avons vu dans les tableaux de la section 5.3 que le comportement des méthodes en présence était similaire dans le cas des chaînes déjà fortement bruitées. Afin de les départager nous faisons tendre le rapport signal sur bruit vers 0 (Figure 17, 18).

Nous constatons la moins bonne compétitivité de l'algorithme MSEM vis-à-vis de deux autres méthodes, lorsque le bruitage devient très important. Notons l'excellent comportement des algorithmes EM et MICE dont le taux de bonne classification suit le taux théorique jusqu'aux niveaux très élevés de bruitage. Précisons à titre d'exemple que pour  $\Delta m = 0.3$  (cas MD) le taux théorique de classification aveugle est de 44%. Ainsi, les résultats des Figures 17 et 18 montrent, d'une part, l'intérêt de la modélisation markovienne et, d'autre part, l'excellence de comportement des algorithmes EM et MICE dans ce contexte.

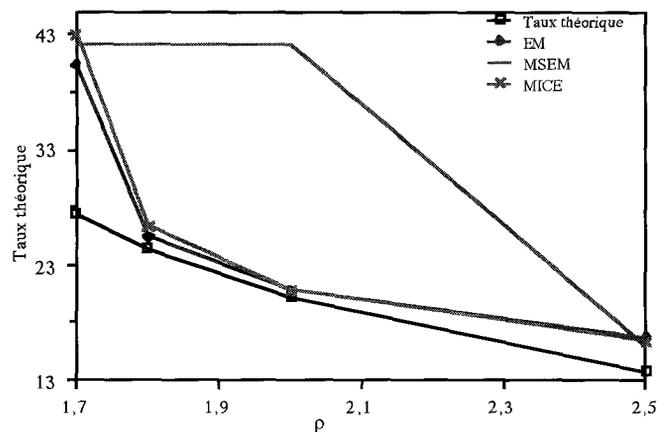


Figure 18. – Taux d'erreur de restauration non supervisée (cas VD) en fonction de  $\rho$  (chaîne A)

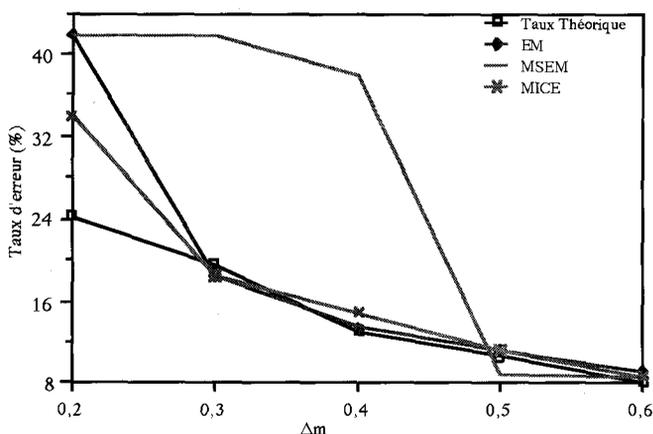


Figure 17. – Taux d'erreur de restauration non supervisée (cas VD) en fonction de  $\Delta m$  (chaîne A)

Par ailleurs, les résultats de la Figure 17 montrent qu'il existe une plage de bruitage ( $0.2 \leq \Delta m \leq 0.3$ ) où MICE est plus efficace que EM.

En conclusion, nous devons opter pour l'utilisation de MICE dans les cas de niveau très élevé de bruitage.

## 7. Segmentation non supervisée d'images

Nous abordons dans cette section le problème de la segmentation statistique non supervisée d'images. La modélisation par champs de Markov cachés est celle généralement retenue par les auteurs. Nous proposons dans la suite une modélisation utilisant les chaînes de Markov cachées, dont l'avantage par rapport à la modélisation précédente se situe à deux niveaux :

1. Rapidité des algorithmes.

Dans les deux types de modélisation la connaissance des différentes lois liées à la loi de  $X$  a posteriori est cruciale, aussi bien au niveau de l'estimation des paramètres que de la segmentation proprement dite. Dans le cas des modélisations par champs de Markov cachés, cette connaissance est acquise par le biais des méthodes itératives, de type de l'échantillonneur de Gibbs, alors qu'elle est obtenue directement, par un calcul analytique, dans celui des modélisations par chaînes de Markov cachées. Il en résulte une différence significative de temps calcul en faveur des modélisations par chaînes de Markov cachées.

2. Souplesse d'utilisation en segmentation spatio-temporelle.

Ainsi que nous le verrons dans la section suivante, l'utilisation des chaînes de Markov cachées permet la conception de nombreux algorithmes de segmentation spatio-temporelle dont la complexité est inférieure à celle des méthodes utilisant les champs de Markov.

La moins bonne adéquation du modèle à la réalité constitue cependant un désavantage de notre méthode. Nous montrons à la fin de cette section que la perte de l'efficacité due à cette moins bonne adéquation semble acceptable. Cependant, des études complémentaires sont nécessaires afin d'évaluer le degré de généralité du phénomène.

**7.1. PARCOURS DE PEANO**

Lorsqu'on modélise l'image par une chaîne de Markov, l'idée la plus simple est de considérer l'image « ligne par ligne » ou « colonne par colonne ». L'inconvénient de cette démarche est que le « passé » et le « futur » d'un pixel dans la chaîne ainsi obtenue ne correspond pas toujours à son contexte spatial de façon satisfaisante. Ainsi, lorsqu'on considère la démarche « ligne par ligne », deux pixels voisins et appartenant à la même colonne sont « proches » spatialement et « éloignés » au sens de la chaîne de Markov. Afin d'améliorer l'adéquation du contexte « temporel » de la chaîne au contexte « spatial » il est possible d'utiliser un parcours de type « Hilbert-Peano » représenté sur la Figure (20) [1], [5], [19]. Ce parcours est obtenu en reproduisant un élément de base, dit « générateur », comme indiqué sur la Figure 19. Notons que l'introduction d'un tel parcours n'est possible que dans le cas des images de taille  $2^n \times 2^n$  auquel nous limitons notre étude. Il existe cependant des parcours similaires permettant de traiter des images de taille quelconque [41].

Conformément à la Figure 20, le premier instant de la chaîne de Markov correspond au pixel de la première ligne et dernière colonne, le deuxième instant correspond au pixel de la deuxième ligne et dernière colonne, le troisième instant correspond au pixel de la deuxième ligne et avant-dernière colonne et ainsi de suite.

Ainsi, en considérant un parcours de ce type dans l'ensemble de pixels, nous obtenons une suite que nous considérerons comme une réalisation bruitée d'une chaîne de Markov. Les algorithmes

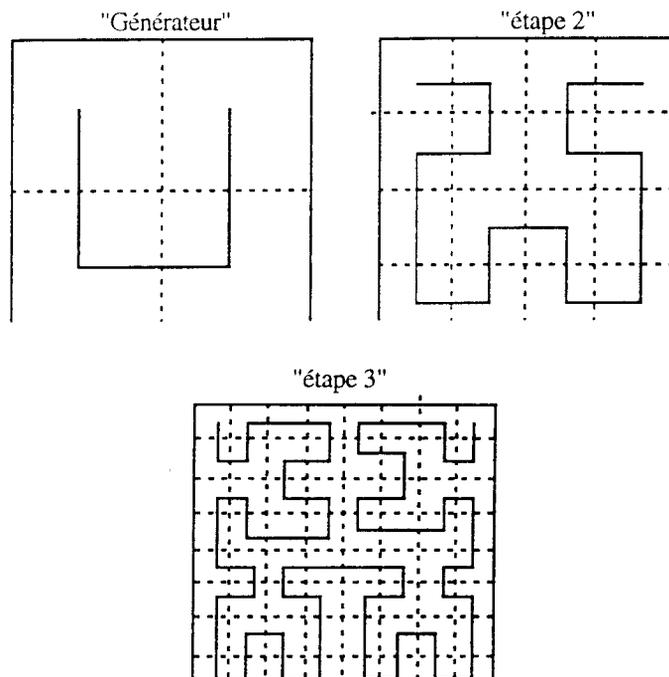


Figure 19.

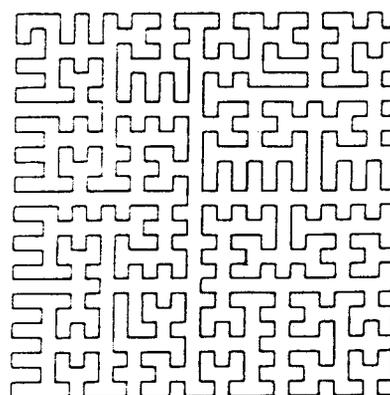


Figure 20.

de segmentation non supervisée sont alors obtenus par l'application à cette suite des algorithmes de restauration non supervisée de la section précédente.

Afin d'étudier le comportement des algorithmes en présence, nous les avons appliqués à quatre images de synthèse « Lettre B », « Anneau », « Gibbs », « Alphabet », d'homogénéités différentes, bruitées avec les 2 types de bruits gaussiens indépendants suivants :

Le bruit MD :  $(\mu_1 = 110, \sigma_1^2 = 900), (\mu_2 = 140, \sigma_2^2 = 900)$

Le bruit VD :  $(\mu_1 = 125, \sigma_1^2 = 225), (\mu_2 = 125, \sigma_2^2 = 900)$

Les images étudiées ainsi que certaines de leurs versions bruitées sont présentées sur les Figures 21, 22.

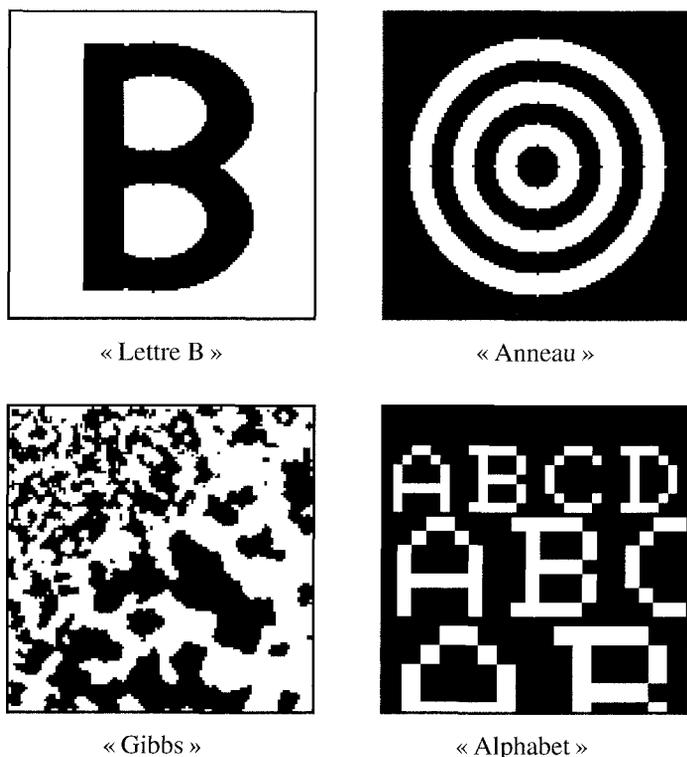


Figure 21.

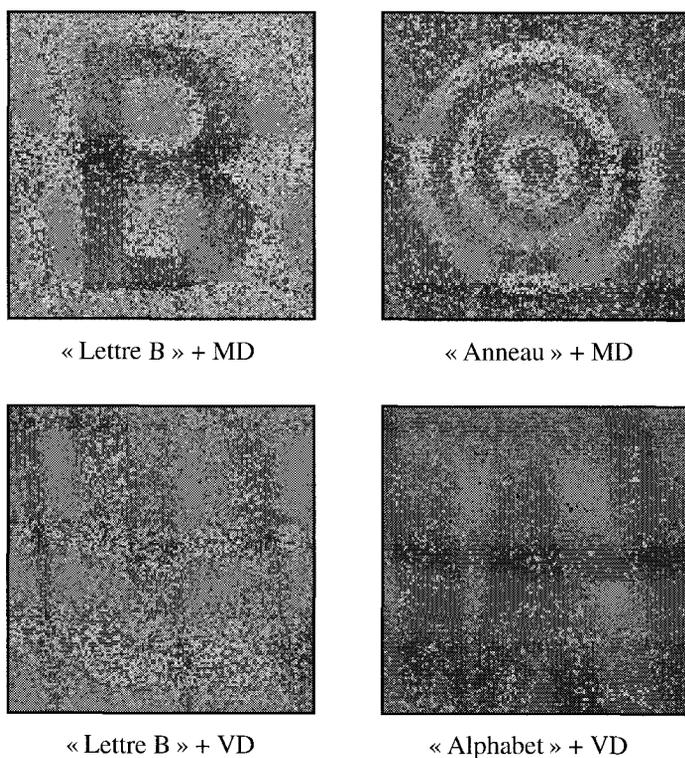


Figure 22.

Nous testons dans un premier temps l'apport de notre démarche par rapport à celle « ligne par ligne ». Le problème de l'estimation

des paramètres d'une chaîne de Markov cachée n'est pas abordé ici; les paramètres de la chaîne sont estimés à partir de l'image non bruitée et les paramètres des bruits sont connus. Ces différents paramètres sont ensuite utilisés dans la segmentation des images par MPM. Les résultats obtenus sont présentés dans le Tableau 7.

Tableau 7. — Calcul des taux d'erreur théoriques (%) pour la segmentation par MPM avec Peano (A-P) et sans Peano (S-P) pour les 4 images

$\tau_{\text{théo}}$ (%)	Lettre « B »		Agneaux		Gibbs		Alphabet	
	bruit MD	bruit VD	bruit MD	bruit VD	bruit MD	bruit VD	bruit MD	bruit VD
S-P	5.41%	6.47%	12.35%	13.31%	20.26%	23.22%	13.53%	13.17%
A-P	3.51%	3.79%	8.25%	8.76%	16.78%	19.19%	5.9%	6.94%

Nous notons une amélioration systématique du taux de pixels bien classés en utilisant la courbe de Hilbert-Peano.

## 7.2. EXPÉRIMENTATIONS

Nous présentons dans les Tableaux 8–11 certains résultats de l'estimation des paramètres et quelques résultats visuels de la segmentation non supervisée par MPM sont donnés par la Figure 23.

Tableau 8. — Segmentation statistique non supervisée de l'image « Lettre B » bruitée par le bruit MD

	VE	EM	MSEM	MICE
$\mu_1$	108.92	110.05	109.88	109.90
$\mu_2$	139.79	139.46	139.47	139.46
$\sigma_1^2$	904.01	931.25	929.14	920.50
$\sigma_2^2$	900.86	914.04	912.46	916.62
$\pi_1$	0.299	0.303	0.301	0.302
$a_{11}$	0.963	0.993	0.989	0.993
$a_{22}$	0.984	0.997	0.997	0.997
$T$	/	36	34	33
$\tau(\%)$	3.51	3.83	3.73	3.78

**Tableau 9. — Segmentation statistique non supervisée de l'image « Lettre B » bruitée par le bruit VD**

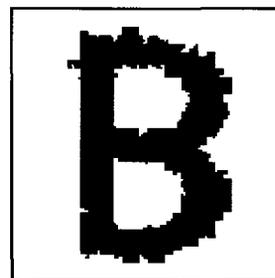
	VE	EM	MSEM	MICE
$\mu_1$	124.52	124.59	124.64	124.61
$\mu_2$	124.59	124.56	124.54	124.55
$\sigma_1^2$	899.02	867.95	867.00	869.02
$\sigma_2^2$	227.99	232.16	232.48	232.19
$\pi_1$	0.299	0.309	0.310	0.309
$a_{11}$	0.963	0.993	0.995	0.994
$a_{22}$	0.984	0.997	0.997	0.997
$T$	/	58	54	56
$\tau(\%)$	3.79	3.86	3.87	3.78

**Tableau 11. — Segmentation statistique non supervisée de l'image « Gibbs B » bruitée par le bruit VD**

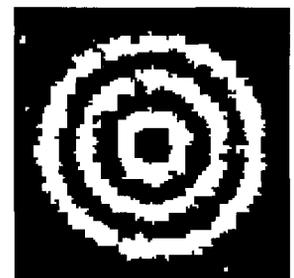
	VE	EM	MSEM	MICE
$\mu_1$	124.59	124.80	124.74	124.92
$\mu_2$	124.91	124.75	124.80	124.65
$\sigma_1^2$	225.64	243.86	244.13	245.09
$\sigma_2^2$	888.79	845.36	946.57	849.06
$\pi_1$	0.459	0.433	0.435	0.437
$a_{11}$	0.873	0.964	0.965	0.965
$a_{22}$	0.892	0.972	0.973	0.973
$T$	/	118	116	113
$\tau(\%)$	19.19	17.84	17.84	17.88

**Tableau 10. — Segmentation statistique non supervisée de l'image « Gibbs B » bruitée par le bruit MD**

VE	EM	MSEM	MICE
109.32	112.53	112.41	112.26
139.72	138.47	138.20	138.62
920.72	993.67	986.37	997.70
911.55	961.82	973.07	947.64
0.459	0.490	0.483	0.492
0.873	0.971	0.975	0.972
0.892	0.972	0.974	0.973
/	58	72	200
16.78	15.78	15.84	15.75



« Lettre B » + MD  
segmentée par  
MICE + MPM



« Anneau » + MD  
segmentée par  
MICE + MPM



« Gibbs B » + VD  
segmentée par  
MICE + MPM



« Alphabet » + VD  
segmentée par  
MICE + MPM

Les résultats des Tableaux 8–11 montrent que la remarquable stabilité des algorithmes observée dans le cas de chaînes de Markov simulées est préservée. Cette stabilité est, a priori, assez surprenante. En effet, le caractère markovien des images n'est nullement acquis et les trois algorithmes d'estimation sont très différents dans leur principe. Par ailleurs, les taux d'erreur obtenus (Tableau 12) sont très proches de ceux obtenus à partir des valeurs empiriques des paramètres, ce qui montre une excellente qualité

**Figure 23.**

de collaboration entre les phases « estimation des paramètres » et « segmentation ».

### 7.3. COMPARAISON AVEC L'ALGORITHME DE VITERBI NON SUPERVISÉ ET L'ALGORITHME EM GIBBSIEN

L'algorithme de Viterbi non supervisé est obtenu par une démarche du type « segmentation courante » [34]. C'est une méthode itérative qui ressemble à l'algorithme SEM : à chaque itération on considère une réalisation de  $X$  obtenue « artificiellement » à partir de la valeur courante des paramètres. Ladite réalisation est obtenue par tirage aléatoire dans le cas de SEM et elle l'est par MAP dans le cas de l'algorithme de Viterbi non supervisé. Elle est ensuite exploitée de façon identique afin d'obtenir la valeur suivante des paramètres. À la stabilisation de la suite des paramètres ainsi obtenue, on procède à la segmentation par le MAP.

Nous présentons ci-dessous la récapitulation des résultats en y adjoignant ceux obtenus par l'algorithme de Viterbi non supervisé [2] et l'algorithme EM Gibbsien (EMG [15]).

**Tableau 12.** — Tableau récapitulatif des taux d'erreur des images avec la segmentation par l'algorithme Viterbi et EM Gibbsien, EM+MPM, MICE+MPM et MSEM+MPM

(%)	Lettre B		Gibbs		Anneau		Alphabet	
	Bruit MD	Bruit VD						
EM	3.83	3.86	15.78	17.84	7.53	8.48	5.82	6.98
MICE	3.78	3.86	15.75	17.88	7.52	8.49	7.79	6.96
MSEM	3.73	3.87	15.84	17.84	7.48	8.44	5.82	6.96
Viterbi	4.14	29.31	20.54	49.73	8.27	50.37	7.99	49.79
EMG	1.78	2.35	18.56	18.00	4.89	6.37	7.09	7.65

Quelques exemples de segmentations par EMG et Viterbi sont donnés sur les Figures 24–27.

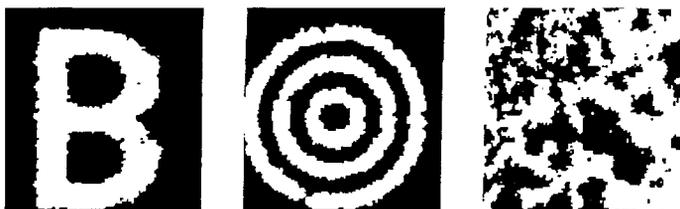


Figure 24. — Les résultats de segmentation de « Lettre B » + MD, « Anneau » + MD, « Gibbs » + MD par EMG

Nous constatons que l'algorithme de Viterbi non supervisé est inefficace dans le cas de bruitage de type VD et n'est pas compétitif vis-à-vis de MICE, MSEM et EM dans le cas de bruitage de type MD. Ainsi le type de bruit (MD ou VD) a une influence déterminante sur cet algorithme. Notons que la mauvaise

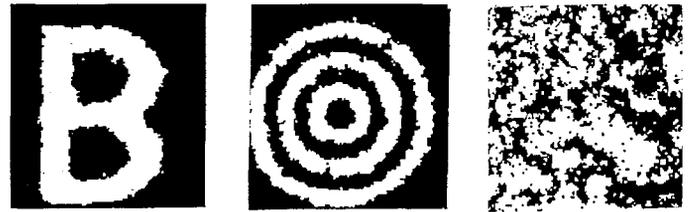


Figure 25. — Les résultats de segmentation de « Lettre B » + MD, « Anneau » + MD, « Gibbs » + MD par Viterbi

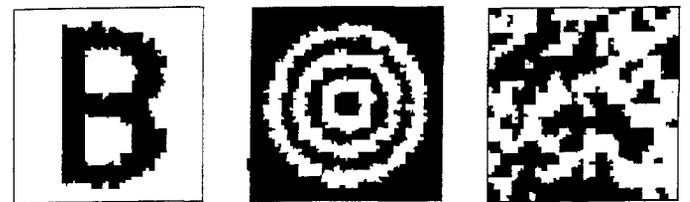


Figure 26. — Les résultats de segmentation de « Lettre B » + MD, « Anneau » + MD, « Gibbs » + MD par Viterbi

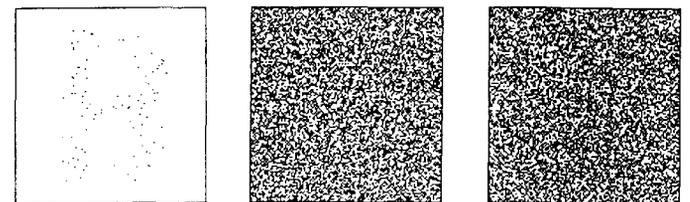


Figure 27. — Les résultats de segmentation de « Lettre B » + VD, « Anneau » + VD, « Gibbs » + VD par Viterbi

qualité de ces résultats peut être due à la phase « estimation des paramètres » et ne met pas en cause l'intérêt de l'algorithme de Viterbi donnant la solution du MAP lorsque les paramètres sont connus.

L'algorithme EM Gibbsien est plus performant dans les cas des images « Lettre B » et « Anneau » par contre, la qualité des méthodes MICE + MPM, MSEM + MPM et EM + MPM est légèrement supérieure dans les cas des images « Gibbs » et « Alphabet ». Le calcul du taux moyen sur les huit cas étudiés pour chacun des algorithmes aboutit au classement suivant : EMG (8.34%), MSEM (9.74%), MICE (8.75%), EM (8.77%), Viterbi (27.52%). Une étude plus approfondie est nécessaire afin de mieux situer les différences de comportement de deux familles de méthodes, toutefois, la perte de la qualité des méthodes MICE + MPM, MSEM + MPM et EM + MPM par rapport à celle de l'algorithme EM Gibbsien apparaît comme acceptable.

Notons cependant que l'EMG utilisé ici est fondé sur une modélisation markovienne la plus simple (champs de Markov relativement aux quatre plus proches voisins); l'utilisation d'un modèle plus perfectionné pourrait éventuellement conduire à un algorithme de type EMG plus performant.

## 8. Segmentation non supervisée spatio-temporelle d'images

Nous proposons dans cette section une généralisation de notre méthode à la segmentation spatio-temporelle. Notons  $I^1, I^2, \dots, I^m$  une suite d'images à segmenter qui seront considérées, en accord avec la modélisation retenue, comme réalisation d'une suite  $Y^1, Y^2, \dots, Y^m$  de champs aléatoires. Nous considérerons dans la suite deux images successives qui seront notées  $I^1, I^2$  et considérées comme réalisations de deux champs aléatoires  $Y^1, Y^2$ . Le problème consiste en segmentation de l'image  $Y^2 = y^2$  en exploitant l'information contenue dans l'image  $Y^1 = y^1$ . Nous proposons trois algorithmes correspondant à trois niveaux d'exploitation de cette information.

### 8.1. SEGMENTATION À PARTIR DE $Y^2 = y^2$

La démarche la plus simple consiste en segmentation de  $Y^2 = y^2$  par la méthode décrite dans la section précédente, en utilisant l'information contenue dans l'image  $Y^1 = y^1$  uniquement au niveau de l'estimation des paramètres. Notons que les images  $Y^1 = y^1$  et  $Y^2 = y^2$  peuvent être très différentes sans que les paramètres définissant les deux modèles le soient. On peut ainsi utiliser les paramètres de la segmentation de  $Y^1 = y^1$  pour initialiser les procédures d'estimation de ceux nécessaires à la segmentation de  $Y^2 = y^2$  et se contenter de très peu, voire d'une seule itération. Dans la suite nous noterons A1 tout algorithme de ce type.

### 8.2. SEGMENTATION À PARTIR DE $Y^1 = y^1$ ET $Y^2 = y^2$

Nous souhaitons segmenter  $Y^2 = y^2$  à partir de  $Y^1 = y^1$  et  $Y^2 = y^2$ . Il est alors nécessaire, afin de se ramener à la modélisation par chaînes de Markov cachées, de définir un parcours de type « Hilbert-Peano ». Nous proposons le parcours suivant :

Si  $n$  est le nombre de pixels, nous obtenons ainsi une suite de longueur  $2n$  à laquelle nous pouvons appliquer les démarches de la section 7. Notons que, de la même manière que dans la section 8.1, nous pouvons utiliser les paramètres de la segmentation de  $Y^1 = y^1$  (à partir de  $Y^1 = y^1$  et l'image précédente) pour initialiser les procédures d'estimation des paramètres nécessaires à la segmentation de  $Y^2 = y^2$  (à partir de  $Y^1 = y^1$  et  $Y^2 = y^2$ ). Cet algorithme sera appelé A2 dans la suite.

### 8.3. DIFFÉRENCIATION DES TRANSITIONS TEMPORELLE ET SPATIALE

Nous reprenons la démarche de la sous-section précédente. Notons  $Y_1^1, Y_2^1, \dots, Y_n^1$  les variables aléatoires correspondant au

parcours de Peano dans l'image  $Y^1 = y^1$  (chaîne à l'instant 1, Figure 28) et  $Y_1^2, Y_2^2, \dots, Y_n^2$  les variables aléatoires correspondant au parcours de Peano dans l'image  $Y^2 = y^2$  (chaîne à l'instant 2, Figure 28). La suite de longueur  $2n$  de la sous-section précédente s'écrit :  $Y_1^2, Y_2^2, Y_2^1, Y_3^1, Y_3^2, \dots, Y_{n-1}^1, Y_{n-1}^2, Y_n^2$  (si  $n$  est paire) et  $Y_1^2, Y_2^2, Y_2^1, Y_3^1, Y_3^2, \dots, Y_n^1, Y_n^2$  (si  $n$  est impaire). Supposons  $n$  paire et retenons la première écriture. Dans la sous-section 8.2 nous avons implicitement supposé que les matrices de transition de la chaîne de Markov  $X_1^2, X_2^2, X_2^1, X_3^1, X_3^2, \dots, X_{n-1}^1, X_{n-1}^2, X_n^2$  correspondante étaient indépendantes du pixel considéré. Une telle hypothèse pourrait s'avérer trop forte. En effet, il existe dans cette chaîne deux transitions de nature différente. Les transitions correspondant aux couples de variables  $X_i^1, X_i^2$  et  $X_i^2, X_i^1$  sont de nature « temporelle » (le même pixel est considéré à deux instants différents dans la séquence), alors que celles correspondant aux couples de variables  $X_i^1, X_{i+1}^1$  et  $X_i^2, X_{i+1}^2$  sont de nature « spatiale » (deux pixels voisins sont considérés à un instant donné dans la séquence). Nous introduisons ainsi deux matrices de transition différentes. L'adaptation des procédures d'estimation et de segmentation décrites dans ce qui précède à ce cas ne pose pas de problèmes particuliers, l'algorithme obtenu sera appelé A3.

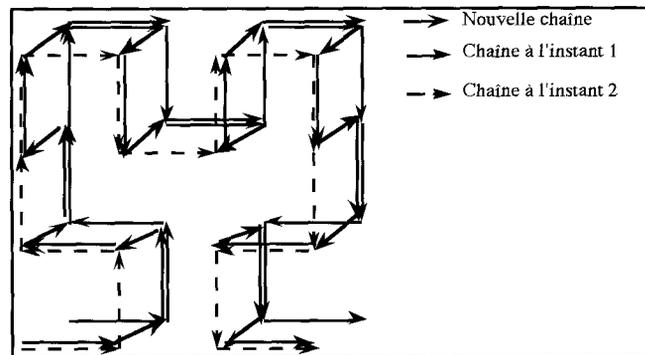


Figure 28.

### 8.4. SEGMENTATION D'UNE SÉQUENCE D'IMAGES RÉELLES

Nous considérons une séquence d'images vidéo réelles et présentons les résultats de sa segmentation par les algorithmes A1, A2, A3 respectivement (Figure 29). Chacune des séquences a été initialisée avec 12 classes. Par un mécanisme de fusion de classes décrit dans [5], les algorithmes se sont stabilisés à 8 classes. Le critère d'arrêt de la section 5.2 est vérifié au bout d'une trentaine d'itérations pour la première image et nécessite 5 à 10 itérations pour les images suivantes.

Nous constatons une nette différence de qualité entre A1 et A2. Cela signifie que l'information contenue dans l'image précédente constitue un apport non négligeable. Notons que l'exploitation de cette information par des techniques utilisant la modélisation par

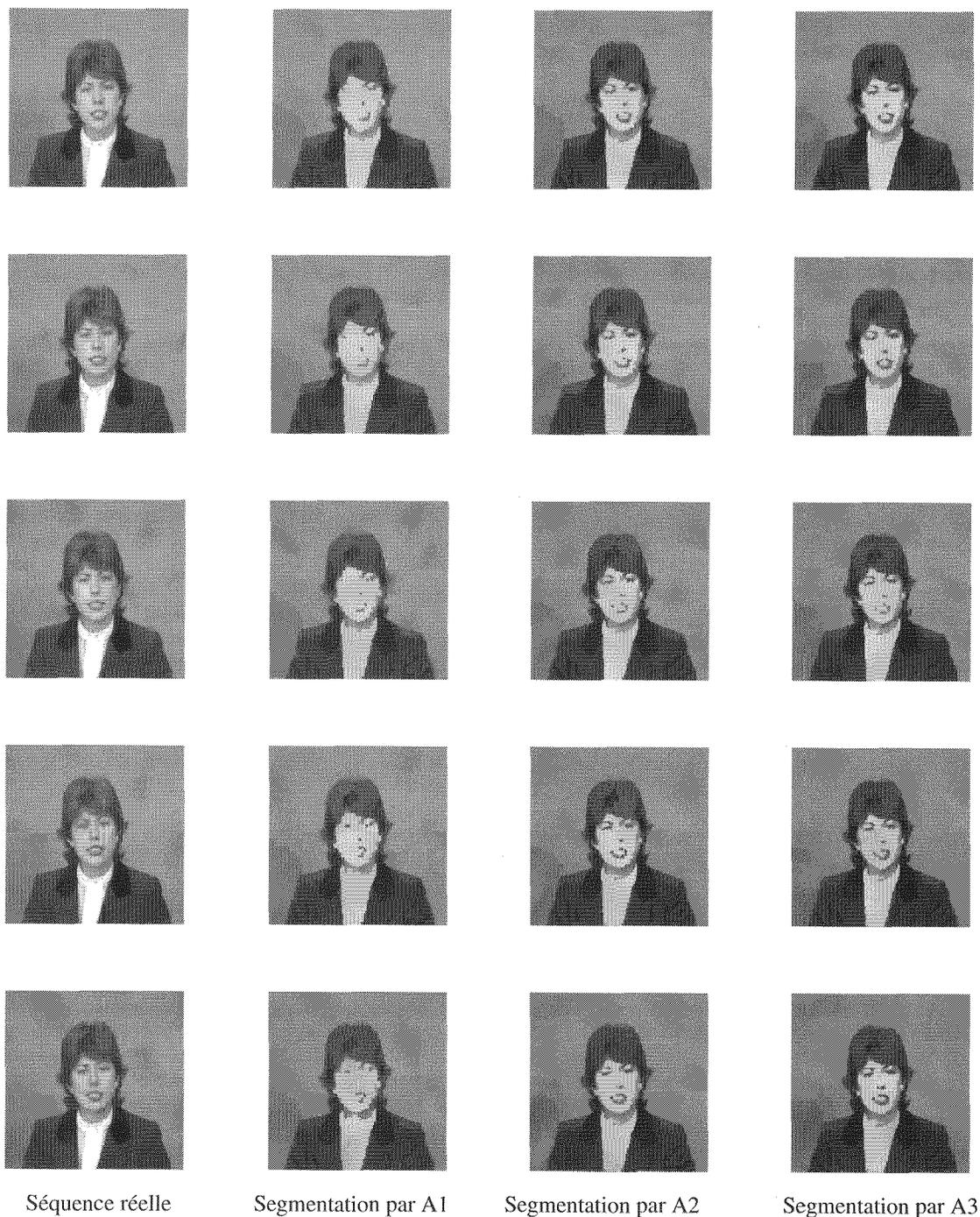


Figure 29.

champs de Markov cachés est possible mais les calculs deviennent rapidement très lourds.

La qualité de la séquence des résultats de segmentations par A3 est meilleure que celle de la séquence des résultats de segmentations par A2, cependant la différence est moins nette.

**Remarque 9**

Dans l'algorithme A3, chaque pixel est classé à partir de l'image courante et de l'image précédente. Notons que la souplesse de l'utilisation de la modélisation par chaînes de Markov cachées et les parcours de Peano autorise un grand nombre de possibilités

de conception d'algorithmes différents. On peut imaginer, à titre d'exemple, que l'essentiel de l'information est concentré dans un cube de côtés  $2r + 1$  centré sur le pixel donné (on considère  $r$  images précédentes et  $r$  images suivantes) et utiliser un parcours de Peano spatial contenant tous les pixels du cube. On peut également imaginer des pyramides dont les bases, centrées sur le pixel donné, ferait partie de l'image courante et dont les sommets seraient orientés soit vers le passé, soit vers le futur. De façon plus générale, en choisissant divers parcours dans l'espace on peut aisément mettre en œuvre autant d'algorithmes de segmentation non supervisée d'images 3D.

#### Remarque 10

Les méthodes d'estimation exposées peuvent être généralisées au cas non stationnaire (la matrice de transition  $A_t$  dépend de l'instant  $t$ ) de la façon suivante : la re-estimation de est faite à partir d'une fenêtre centrée sur  $t$ . Les méthodes de ce type sont beaucoup plus coûteuses en temps calcul et les résultats numériques que nous avons obtenus s'avèrent peu concluants [6]. Notons cependant que ce type de démarche apporte des améliorations de résultats parfois spectaculaires dans le contexte de la segmentation non supervisée locale [31], [36] et il est probable que des études plus approfondies mettraient en lumière des situations où ces méthodes présentent un intérêt.

## 9. Conclusion

Nous avons étudié dans ce travail le problème de l'estimation des paramètres dans les chaînes de Markov cachées et les possibilités des applications des solutions proposées au problème de la segmentation statistique non supervisée d'images. Deux méthodes originales d'estimation, obtenues par des modifications des méthodes générales Stochastic Estimation Maximisation (SEM) et Iterative Conditional Estimation (ICE), ont été proposées. L'étude numérique comparative de ces méthodes, appelées MSEM et MICE respectivement, et de la méthode Expectation-Maximisation (EM) a permis de mettre en évidence leur bon comportement dans diverses situations d'homogénéité et de bruitage des chaînes. Toutefois, lorsque le rapport signal sur bruit tend vers 0, MSEM devient non compétitif vis-à-vis de MICE et EM. Par ailleurs, dans certaines situations particulières, les performances de MICE sont supérieures à celle de EM. L'adjonction à ces algorithmes de la méthode bayésienne Mode de la Marginale a Posteriori (MPM) de restauration permet la conception de trois algorithmes de restauration non supervisée. Le comportement de ces algorithmes est excellent : leurs taux de bonne classification sont très proches, même dans le cas de bruitages importants, de celui obtenu par MPM utilisant les paramètres estimés à partir des données complètes. Tout se passe comme si le fait que la chaîne est inobservable n'avait pas d'influence sur le taux de bonne classification.

L'utilisation de parcours de type Peano dans l'ensemble de pixels permet la conception, à partir des méthodes EM+MPM, MICE+MPM, MSEM+MPM mentionnées ci-dessus, de trois méthodes de segmentation statistique non supervisée d'images. La rapidité et la souplesse de ces méthodes constituent un avantage par rapport aux algorithmes fondés sur des modélisations par champs de Markov cachés. Cependant, l'adéquation du modèle sous-jacent à la réalité est moins bonne. Divers tests portant sur des images de synthèse montrent le bon comportement de ces méthodes, par ailleurs, la perte de l'efficacité due à la moins bonne adéquation du modèle à la réalité semble acceptable.

La souplesse de la modélisation proposée permet la conception de plusieurs algorithmes de segmentation statistique spatio-temporelle non supervisée. Nous en avons proposé trois, de degré de perfectionnement croissant, et avons présenté les résultats de leur application à la segmentation d'une séquence d'images réelles.

En conclusion, la rapidité et la souplesse des méthodes utilisant les chaînes de Markov cachées, les parcours de type Hilbert-Peano, l'un des algorithmes EM, MICE, MSEM pour l'estimation des paramètres et la méthode MPM pour la segmentation, semblent ouvrir des perspectives originales au traitement du problème de la segmentation statistique spatio-temporelle non supervisée d'images. De façon plus générale, ces techniques sont applicables aux problèmes de segmentation non supervisée des images 3D, voire des séquences d'images 3D.

*Ce travail a été partiellement financé par la convention CNET-INT 93 PE 74 05. Les auteurs remercient Bruno Choquet du CCETT Rennes et Olivier Avaro du CNET Paris A pour leur collaboration.*

## BIBLIOGRAPHIE

- [1] K. Abend, T.J. Harley, L.N. Kanal, « Classification of binary random patterns », *IEEE Transactions on Information Theory*, Vol. IT-11, N° 4, 1965.
- [2] M. Avila, C. Olivier, T. Paquet, Y. Lecoutier, « Procédure de reconnaissance de l'écriture manuscrite, basée sur des chaînes de Markov cachées, et appliquée à un vocabulaire limité », *Actes de Quatorzième Colloque GRETSI 93*, Juans-les-Pins, France, 1993, pp. 803-806.
- [3] Azencott R. Ed., « Simulated annealing : Parallelization techniques », Wiley, 1992.
- [4] L.E. Baum, T. Petrie, G. Soules, N. Weiss, « A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains », *Ann. Math. Statistic.*, 41, 1970, pp. 164-171.
- [5] B. Benmiloud, A. Peng, W. Pieczynski, « Estimation conditionnelle itérative dans les chaînes de Markov cachées et segmentation statistique non supervisée d'images », *Actes de Quatorzième Colloque GRETSI 93*, Juans-les-Pins, France, 1993, pp. 105-108.
- [6] B. Benmiloud, « Chaînes de Markov cachées et segmentation statistique non supervisée de séquence d'images », *Thèse*, Université Paris VII, 1994.
- [7] A. Benveniste, M. Métivier, B. Priouret, « Algorithmes adaptatifs et approximations stochastiques », *Techniques stochastiques*, Masson, Paris, 1987.
- [8] J. Besag, « On the statistical analysis of dirty pictures », *Journal of the Royal Statistical Society, Series B*, 48, 1986, pp. 259-302.
- [9] R. Boite, M. Kunt, « Traitement de la parole », *Presses Polytechniques Romandes*, 1987.

- [10] P. Bouthemy, « Modèles et méthodes pour l'analyse du mouvement dans une séquence d'images », *Technique et Science Informatique*, Vol. 7, N°6, 1988, pp. 527-543.
- [11] P. Bouthemy, P. Lalande, « Motion detection in an image sequence using Gibbs distribution », *Actes d'ICASSP'89*, Glasgow, 1989, pp. 1651-1654.
- [12] B. Braathen, W. Pieczynski, P. Masson, « Global and local methods of unsupervised Bayesian segmentation of images », *Machine Graphics and Vision*, Vol. 2, N°1, 1993, pp. 39-52.
- [13] H. Caillol, A. Hillon, W. Pieczynski, « Fuzzy random fields and unsupervised image segmentation », *IEEE Transactions on GRS*, Vol. 31, N°4, 1993, pp. 801-810.
- [14] G. Celeux, J. Diebolt, « L'algorithme SEM : un algorithme d'apprentissage probabiliste pour la reconnaissance de mélanges de densités », *Revue de Statistique Appliquée*, Vol. 34, N°2, 1986.
- [15] B. Chalmond, « An iterative Gibbsian technique for reconstruction of m-ary images », *Pattern Recognition*, Vol. 22, N°6, 1989, pp. 747-761.
- [16] M.M. Dempster, N.M. Laird, D.B. Rubin, « Maximum likelihood from incomplete data via the EM algorithm », *Journal of the Royal Statistical Society, Series B*, 39, 1977, pp. 1-38.
- [17] P.A. Devijver, « Baum's forward-backward algorithm revisited », *Pattern Recognition Letters*, 3, 1985, pp. 369-373.
- [18] P.A. Devijver, M. Dekesel, « Champs aléatoires de Pickard et modélisation d'images digitales », *Traitement du Signal*, Vol. 5, N°5, 1988, pp. 131-150.
- [19] M. Emsalem, H. Caillol, P. Ollivier, G. Carnat, W. Pieczynski, « Fast unsupervised statistical image segmentation », *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 92)*, Houston, Texas, 1992.
- [20] G.D. Fornay, « The Viterbi algorithm », *Proceedings of the IEEE*, Vol. 61, N°3, 1973, pp. 268-277.
- [21] S. Geman and D. Geman, « Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images », *IEEE Transactions on PAMI*, Vol. PAMI-6, N°6, 1984, pp. 721-741.
- [22] X. Guyon, « Champs aléatoires sur un réseau », *Collection Techniques Stochastiques*, Masson, Paris, 1993.
- [23] F. Heitz, P. Bouthemy, « Motion estimation and segmentation using a global bayesian approach », *IEEE Actes de ASSP'90*, Albuquerque, 1990.
- [24] F. Heitz, P. Bouthemy, « Estimation et segmentation du mouvement : approche Bayésienne et modélisation Markovienne des occlusions », *Actes de la 7ème Conf. AFCET / RFA*, Paris, 1989, pp. 1359-1368.
- [25] A. Hillion, « Les approches statistiques pour la reconnaissance des images de télédétection », *Atti della XXXVI Riunione Scientifica, SIS*, Vol. 1, 1992, pp. 287-297.
- [26] S. Lakshmanan, H. Derin, « Simultaneous parameter estimation and segmentation of Gibbs random fields using simulated annealing », *IEEE Transactions on PAMI*, Vol. 11, N°8, 1989, pp. 799-813.
- [27] P. Lalande, « Détection du mouvement dans une séquence d'images selon une approche markovienne; Application à la robotique sous-marine », *Thèse*, Université de Rennes I, 1990.
- [28] N. Marhic, P. Masson, W. Pieczynski, « Mélange de lois et segmentation non supervisée des données SPOT », *Statistique et Analyse des Données*, Vol. 16, N°2, 1991, pp. 59-79.
- [29] J.L. Marroquin, S. Mittle, T. Poggio, « Probabilistic solution of ill-posed problems in computational vision », *Journal of the American Statistical Association*, Vol. 82, 1987, pp. 76-89.
- [30] P. Masson, W. Pieczynski, « SEM algorithm and unsupervised statistical segmentation of satellite images », *IEEE Transactions on GRS*, Vol. 34, N°3, 1993, pp. 618-633.
- [31] A. Peng, W. Pieczynski, « Adaptive mixture estimation and unsupervised contextual Bayesian image segmentation », *Graphical Models and Image Processing*, Vol. 57, N°5, 1995, pp. 389-399.
- [32] W. Pieczynski, « Estimation of context in random fields », *Journal of Applied Statistics*, Vol. 16, N°2, 1989, pp. 283-290.
- [33] W. Pieczynski, « Statistical image segmentation », *Machine Graphics and Vision*, Vol. 1, N°1/2, 1992, pp. 261-268.
- [34] W. Pieczynski, « Champs de Markov cachés et estimation conditionnelle itérative », *Traitement du Signal*, Vol. 11, N°2, 1994, pp. 141-153.
- [35] W. Pieczynski, J.-M. Cahen, « Champs de Markov flous cachés et segmentation d'images », *Revue de Statistique Appliquée*, Vol. 42, N°3, 1994, pp. 13-31.
- [36] H.C. Quelle, J.-M. Boucher, W. Pieczynski, « Local parameter estimation and unsupervised segmentation of SAR images », *Actes de IGARSS'92*, Houston, Texas, 1992.
- [37] W. Qian, D.M. Titterton, « On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions », *Journal of Applied Statistics*, Vol. 16, N°2, 1989, pp. 267-282.
- [38] J. Quinqueton, « Le concept de dimension intrinsèque en reconnaissance des formes », *Thèse*, Université Paris VI, 1981.
- [39] L.R. Rabiner, S.E. Levinson, M.M. Sondhi, « On the use of hidden Markov model for speaker-independent recognition of isolated words from of medium-size vocabulary », *AT&T Bell Laboratory Technical Journal*, 1984, pp. 627-642.
- [40] L.R. Rabiner, « A tutorial on hidden Markov models and selected applications in speech recognition », *Proceedings of IEEE*, Vol. 77, No2, 1989, pp. 257-286.
- [41] W. Skarbek, « Generalized Hilbert scan in image printing », *Theoretical Foundations of Computer Vision*, R. Klette et W. G. Kropetsh Ed., Akademik Verlag, 1992, pp. 45-57.
- [42] J. Tilton, S. Vardeman, P. Swain, « Estimation of context for statistical classification of multispectral image data », *IEEE Transactions on GRS*, GE-20, 1982, pp. 445-452.
- [43] Y. He, A. Kundu, « 2-D Shape Classification Using Hidden Markov Model », *IEEE Transactions on PAMI*, Vol. 13, N°11, 1991, pp. 1172-1184.
- [44] J.F. Yao, « Segmentation bayésienne d'images : comparaisons de méthodes contextuelles et globales », *Cahier du Centre d'Études et de Recherches*, Op.30 (4), 1989, pp. 269-290.
- [45] L. Younes, « Parametric inference for imperfectly observed Gibbsian fields », *Probability Theory and Related Fields*, 82, 1989, pp. 625-645.

Manuscrit reçu le 30 Mai 1994.

## LES AUTEURS

### Btissam BENMILOUD



**Titulaire d'un Diplôme d'Études Approfondies, Université Paris 7, Btissam Benmiloud a effectué ses travaux de recherches de thèse à l'Institut National des Télécommunications d'Évry. Ces derniers portaient sur l'utilisation des modèles Markoviens cachés en segmentation non supervisée de séquences d'images et ont été sanctionnés par le Doctorat de l'Université Paris 7, mention Méthodes Physiques en Télédétection, en 1994.**

### Wojciech PIECZYNSKI



**Titulaire d'un Doctorat d'État en Statistique Mathématique, Université Paris 6, Wojciech Pieczynski a enseigné à l'Université de Tiaret, Algérie, à l'Université de Brazzaville, Congo, et l'École Nationale Supérieure des Télécommunications de Bretagne. Actuellement Professeur à l'Institut National des Télécommunications d'Évry, ses recherches portent sur les approches probabilistes du traitement d'images.**