

Reconnaissance de caractères manuscrits par Réseaux Bayésiens Dynamiques

Khalid Hallouli, Laurence Likforman-Sulem, Marc Sigelle

► **To cite this version:**

| Khalid Hallouli, Laurence Likforman-Sulem, Marc Sigelle. Reconnaissance de caractères manuscrits
| par Réseaux Bayésiens Dynamiques. Jun 2004, 2004. <sic_00001163>

HAL Id: sic_00001163

https://archivesic.ccsd.cnrs.fr/sic_00001163

Submitted on 6 Dec 2004

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

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

Reconnaissance de caractères manuscrits par Réseaux Bayésiens Dynamiques

Khalid Hallouli – Laurence Likforman-Sulem – Marc Sigelle

École Nationale Supérieure des Télécommunications
Département de Traitement de Signal et des Images
46 rue Barrault 75634 Paris Cedex 13 France

hallouli,likforman,sigelle@tsi.enst.fr

Résumé : *Ce travail présente une approche pour la reconnaissance des caractères manuscrits qui a l'avantage de surmonter les limitations des modèles de Markov Cachés (HMMs). L'idée principale consiste à construire des modèles 2D en couplant dans un réseau bayésien dynamique deux réseaux de type HMM. Le premier modèle HMM est obtenu à partir d'observations de type colonnes de pixels (HMM-vertical), le second à partir d'observations de type lignes (HMM-horizontal). L'un des avantages de ces modèles est de nous permettre d'obtenir une bonne modélisation des images des caractères manuscrits, due aux regroupement des informations concernant les lignes et les colonnes. Les résultats expérimentaux, obtenus sur la base de chiffres MNIST montrent que l'approche par réseaux bayésiens est très prometteuse dans le domaine de la reconnaissance de l'écrit.*

Mots-clés :

Intelligence artificielle, Réseau Bayésien Dynamique, Reconnaissance de caractères manuscrits, Inférence

1 Introduction

Les modèles probabilistes, tels que les modèles de Markov cachés (HMMs), sont largement utilisés dans les domaines de la reconnaissance de la parole et de l'écriture, car ils savent bien s'adapter à la variabilité des observations, notamment les distorsions élastiques. Même si ces modèles HMMs ont de bonnes performances, les HMMs sont cependant des modèles essentiellement mono-dimensionnels. Des extensions à l'aspect 2D de l'écriture ont été réalisées avec les modèles pseudo-2D (HMM-Planar) [SAO 97], les modèles à champs de Markov [PAR 98] et la programmation dynamique 2D [GEO 04]. Les premiers supposent cependant l'indépendance entre les lignes et les colonnes des caractères, les seconds utilisent les régions (ensembles de cliques).

En exploitant les corrélations entre variables, une modélisation plus fine peut être obtenue. Dans cette optique, des modèles probabilistes s'appuyant sur les réseaux bayésiens sont récemment apparus dans le domaine du traitement de l'écrit appliqués à la reconnaissance de caractères [CHO 03] et l'analyse de document [SOU 02].

Notre approche est basée sur le couplage de deux modèles de type HMM, en faisant apparaître des liens supplémentaires entre variables (cachées et observables). Pour cela nous uti-

lisons le formalisme des Réseaux Bayésiens (RB). Ce dernier est devenu un formalisme puissant qui unifie différents concepts de modélisation probabiliste utilisés en statistique, intelligence artificielle et reconnaissance de la parole. Ce formalisme est couramment employé pour le raisonnement sous incertain et peut être considéré comme un système expert probabiliste. Les réseaux bayésiens constituent un cadre intéressant pour deux raisons. D'une part, grâce à leurs structures graphiques, les RBs offrent un outil naturel pour représenter les dépendances entre les différentes variables d'un système donné. D'autre part, en exploitant les indépendances conditionnelles entre les variables, ils introduisent une certaine modularité dans les systèmes complexes.

Nous modélisons les dépendances entre les nœuds (variables aléatoires) en créant des interactions entre deux réseaux : un HMM-horizontal et un HMM-vertical. Le réseau HMM-vertical (resp. HMM-horizontal) prend pour séquence d'entrée (observation) les colonnes de pixels du caractère (resp. les lignes de pixels) [HAL 02]. La modélisation 2D résulte du couplage entre variables d'états ou variables d'observations de ces deux réseaux. Ainsi notre approche est bien une modélisation bidimensionnelle puisqu'elle regroupe les données selon les lignes et les colonnes simultanément.

Nous rappelons tout d'abord le formalisme des réseaux bayésiens dans le cas statique et dynamique et le mécanisme d'inférence puis nous présentons les modèles de couplage et nos résultats sur la base de caractères manuscrits MNIST qui montrent la pertinence de cette approche.

2 Réseaux Bayésiens

Un réseau bayésien (réseau probabiliste ou Bayesian Networks BNs) [JEN 88, LAU 99] est un modèle représentant des connaissances incertaines sur un phénomène complexe, et permettant à partir des données un véritable raisonnement. Ainsi un réseau bayésien statique est un graphe acyclique orienté dont les nœuds sont des variables aléatoires qui ont des valeurs discrètes, ou continues selon une loi Gaussienne. Il est constitué de deux composantes :

- une structure qui est un graphe causal, G orienté, acyclique, dont les nœuds sont des variables d'intérêt du domaine, les arcs des relations de dépendance entre ces variables. L'ensemble des nœuds et des arcs forme ce que l'on appelle la structure graphique du réseau bayésien : c'est la représentation qualitative de la connaissance.

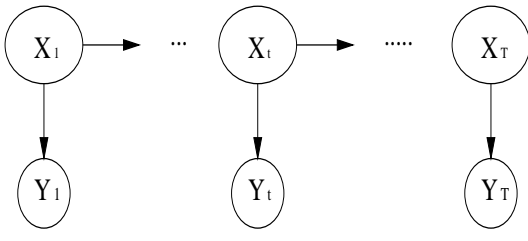


FIG. 1 – HMM représenté comme un RBD avec les $(X_t)_{1 \leq t \leq T}$ représentant les états cachés et les $(Y_t)_{1 \leq t \leq T}$ représentant les observations

- Une paramétrisation θ qui est un ensemble de distributions locales de probabilités conditionnelles (CPD). Pour chaque nœud on dispose d'une table de probabilités conditionnelles $P(\text{variable}/\text{parents}(\text{variable}))$ qui représente la distribution locale de probabilité. Il faut remarquer que l'état de chaque nœud ne dépend que de l'état de ses parents : c'est la représentation quantitative de la connaissance.

Soient un ensemble de variable aléatoires $X = (X_1, X_2, \dots, X_N)$ et $P(X)$ sa distribution jointe de probabilité. La probabilité jointe peut être écrite sous la forme

$$P(X_1, X_2, X_3, \dots, X_n) = \prod_{1 \leq i \leq n} P(X_i | C(X_i)) \quad (1)$$

où $C(X_i)$ est l'ensemble des causes (parents) de X_i dans le graphe G .

La sémantique des indépendances conditionnelles d'un réseau bayésien implique qu'une variable est indépendante de toutes les autres variables du réseau connaissant ses parents, à l'exception de ses descendants.

Les réseaux bayésiens dynamiques (RBD) [MUR 02] sont une extension des réseaux bayésiens qui représente l'évolution temporelle des variables aléatoires. On considère un ensemble $X_t = (X_t^1, \dots, X_t^N)$ de variables évoluant dans $[1, T]$. La probabilité jointe peut être codée par un réseau statique avec $T * N$ variables avec une possibilité d'avoir une structure ou des paramètres différents à chaque instant t . Si le processus est stationnaire, les hypothèses d'indépendance et les probabilités conditionnelles associées sont identiques pour tous les temps t . Dans ce cas un RBD peut être représenté par un réseau bayésien dont la structure est répétée à chaque instant t . Ainsi un HMM est un cas particulier d'un RBD (voir Fig.1). Dans le cas où les vecteurs d'observations suivent une distribution gaussienne, les dépendances entre variables d'états discrètes \mathbf{X} et les observations \mathbf{Y} s'expriment par :

$$P(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{M})^t \Sigma^{-1} (\mathbf{y} - \mathbf{M})\right] \quad (2)$$

où

- d est la dimension de l'espace de représentation.
- \mathbf{M} est le vecteur moyenne.
- Σ est la matrice de covariance.

3 Inférence dans les réseaux Bayésiens

L'utilisation essentielle des réseaux bayésiens est de calculer des probabilités conditionnelles d'événements reliés les uns aux autres par des relations de cause à effet et que l'on appelle *inférence*.

Supposons que le graphe soit constitué de n nœuds notés $X = X_1, X_2, \dots, X_n$. Le problème général de l'inférence consiste à calculer

$$P(X_i | Y) \quad \text{où } Y \subset X, \quad X_i \notin Y$$

Souvent Y est l'ensemble des variables ou vecteurs d'observations noté O .

Pour calculer ces probabilités conditionnelles on peut utiliser des algorithmes exacts ou approchés. L'algorithme le plus utilisé est l'algorithme de l'arbre de jonction [ZWE 98] qui est similaire à l'algorithme de Baum-Welch utilisé dans les HMMs. Pour cela on transforme le réseau initial en une nouvelle structure appelée *arbre de jonction*. Cet arbre s'obtient en suivant les étapes suivantes :

Moralisation : en reliant les parents entre eux et en éliminant les directions.

Triangularisation : en ajoutant sélectivement des arcs au graphe moral (pour ne pas avoir des cycles d'ordre quatre ou plus).

Arbre de jonction : est obtenu à partir du graphe triangulé en connectant les cliques de telle façon que toutes les cliques sur le chemin entre deux cliques X et Y contiennent $X \cap Y$.

Nous rappelons qu'une *clique* est un sous graphe complet maximal :

- complet : chaque paire de nœuds (variables) distincts est connectée par un arc.
- maximal : la clique n'est pas strictement contenue dans un sous graphe complet.

Sur un arbre, l'inférence se base sur le calcul de deux variables λ et π :

$$\lambda_j^i = P(O_i^0, O_i^- | X_i = j) \quad (3)$$

$$\pi_j^i = P(O_i^+, X_i = j) \quad (4)$$

où

- O_i^0 est la valeur observée de X_i dans le cas où X_i est observable
- O_i^- est l'ensemble des observations strictement en aval de X_i
- O_i^+ est l'ensemble des observations strictement en amont de X_i

Ainsi on a

$$\begin{aligned} P(X_i = j, O) &= P(O_i^0, O_i^-, O_i^+, X_i = j) \\ &= P(O_i^+, X_i = j) \times P(O_i^0, O_i^- | X_i = j) \\ &= \lambda_j^i \times \pi_j^i \end{aligned} \quad (5)$$

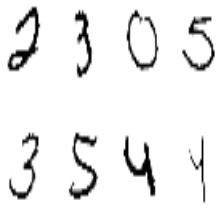


FIG. 2 – Exemples de chiffres (base MNIST)

et donc

$$P(X_i = j | O) = \frac{\lambda_j^i \times \pi_j^i}{\sum_j \lambda_j^i \times \pi_j^i} \quad \forall i \quad (6)$$

$$P(O) = \sum_j \lambda_j^i \times \pi_j^i \quad \forall i \quad (7)$$

Les deux variables λ et π sont analogues aux variables backward et forward α et β utilisées dans les HMMs. Ces variables peuvent se calculer de la façon suivante :

1. Calcul de λ

si X_i est un nœud feuille alors

$$\lambda_j^i = 1 \quad \forall j \quad \text{avec} \quad X_i = j \quad (8)$$

sinon

$$\lambda_j^i = \prod_{f \in \text{fils}(X_i)} \sum_k \lambda_k^f * P(X_f = k | X_i = j) \quad (9)$$

2. Calcul de π

si X_i est le nœud racine alors

$$\pi_j^i = P(X_i = j) \quad (10)$$

sinon

$$\pi_j^i = \sum_v P(X_i = j | X_p = v) * \pi_v^p * \prod_{s \in \text{freres}(X_i)} \sum_k \lambda_k^s * P(X_s = k | X_p = v) \quad (11)$$

avec X_p le parent de X_i dans l'arbre

Les λ_j^i se calculent des feuilles vers la racine de l'arbre, tandis que les π_j^i se calculent de la racine vers les feuilles.

4 Modélisation

Étant donné un chiffre donné (voir Fig.2), notre objectif est qu'au lieu de considérer un seul HMM vertical ou horizontal pour le modéliser, on construit un RBD plus complexe en couplant deux HMMs. Ce couplage est réalisé en ajoutant des liens (orientés) entre les variables (nœuds). Un des problèmes rencontrés est : quels sont les liens à ajouter ? La meilleure solution est d'apprendre la structure graphique à partir des données. Cette stratégie, appelée *apprentissage structurel* [DEV 01], n'est pas l'objectif de notre approche. Notre but est de fixer une structure graphique puis de l'évaluer.

4.1 Structures graphiques utilisées

Nous fixons deux structures en se basant sur les critères suivants :

- le modèle construit doit avoir un nombre raisonnable de paramètres pour que la complexité des calculs reste abordable.
- aucune variable continue ne doit avoir de fils discret afin de pouvoir appliquer l'algorithme d'inférence exacte (arbre de jonction) [JEN 94, LAU 99].
- des liens doivent exister entre les variables cachées.

On note, pour $j = 1, 2$ (X_t^j) $_{1 \leq t \leq T}$ les variables d'états cachées et (Y_t^1) $_{1 \leq t \leq T}$ (resp. (Y_t^2) $_{1 \leq t \leq T}$) les séquences d'observations correspondant aux colonnes des pixels du caractère (resp. lignes) comme le montre la figure 5.

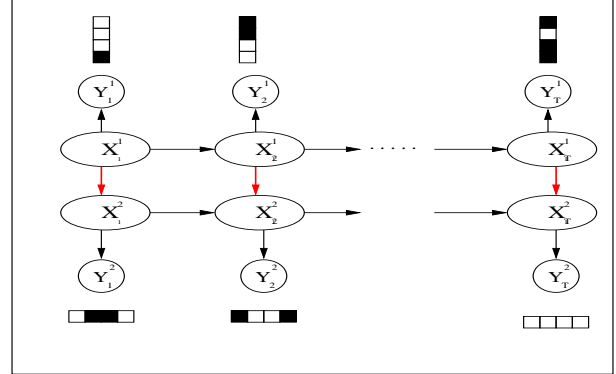


FIG. 3 – Premier modèle de couplage sous forme de RBD : coup1

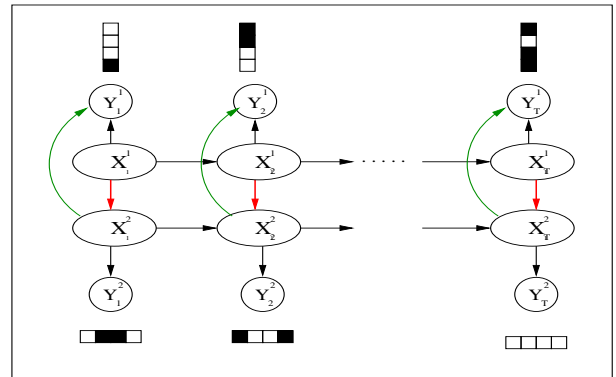


FIG. 4 – Deuxième modèle de couplage sous forme de RBD : coup2

Les figures : Fig.3 et Fig.4 présentent deux modèles de couplage, coup1 et coup2 qui satisfont aux critères annoncés. Les $X_t^j = (X_t^j)_{1 \leq t \leq T}$ pour $j = 1, 2$ sont des variables discrètes prenant leurs valeurs dans un ensemble $I = 1, \dots, k$ avec k le nombre des états cachés, Y_t^1 (respectivement Y_t^2) est une variable continue avec une gaussienne comme distribution représentant l'observation des fichiers des caractères correspondant aux colonnes de pixels (respectivement aux lignes de pixels) à chaque instant t .

coupl1 : est un modèle à base de RBD (Fig.3). Il est construit en couplant les deux HMMs vertical et horizontal en liant les états cachés entre eux et en prenant en compte le fait que le HMM vertical est meilleur que le HMM horizontal [HAL 03]. Ainsi les variables d'états cachées $(X_t^1)_{1 \leq t \leq T}$ sont liées à $(X_t^2)_{1 \leq t \leq T}$. Pour chaque $j \in \{1, 2\}$ les (X_t^j) agissent sur les (Y_t^j) , pour chaque caractère et pour $(i, j, k) \in I^3$ où I est l'espace des états. La paramétrisation numérique θ est définie par la donnée de A, U, b^1 et b^2 :

$$\begin{cases} A_{i,j} = P(X_t^1 | X_{t-1}^1) & t \geq 2 \\ U_{i,j,k} = P(X_t^2 = k | X_{t-1}^2 = i, X_t^1 = j) & t \geq 2 \\ b_i^l(.) = P(Y_t^l = . | X_t^l = i) & \text{pour } l = 1, 2 \end{cases} \quad (12)$$

coupl2 : est modèle à base de RBD (Fig.4), où un lien supplémentaire a été ajouté au premier modèle coupl1, ie., les états X_t^2 contrôlent les deux observations lignes et colonnes (Y_t^1) et (Y_t^2) , pour chaque caractère et pour $(i, j, k) \in I^3$. La paramétrisation numérique θ est définie par la donnée de A, U, b^1 et b^2 :

$$\begin{cases} A_{i,j} = P(X_t^1 | X_{t-1}^1) & t \geq 2 \\ U_{i,j,k} = P(X_t^2 = k | X_{t-1}^2 = i, X_t^1 = j) & t \geq 2 \\ b_{i,j}^1(.) = P(Y_t^1 = . | X_t^1 = i, X_t^2 = j) \\ b_i^2(.) = P(Y_t^2 = . | X_t^2 = i) \end{cases} \quad (13)$$

Ces deux modèles de couplages sont plus riches d'informations puisqu'ils tiennent compte de l'information selon les lignes et les colonnes :

- le premier modèle est un couplage simple en tenant compte du fait que le HMM-vertical est meilleur que le HMM-horizontal.
- le deuxième modèle est plus complexe que le premier (plus de paramètres) et il donne plus d'importance aux observations colonnes qu'aux observations lignes.

4.2 Estimation des paramètres

Dans les expériences que nous effectuons, l'apprentissage des modèles est effectué de façon indépendante modèle par modèle (i.e chiffre par chiffre), on utilise l'algorithme EM (Expectation Maximisation) car les variables d'états sont cachées. L'étape E de l'EM est réalisée en utilisant l'algorithme de Jensen Lauritzen Olesen (JLO) [LAU 99]. Les formules de re-estimation des paramètres peuvent être trouvées dans [DAO 01].

5 Résultats expérimentaux

Pour nos expériences, nous avons utilisé la base de données MNIST qui est constituée des images de chiffres manuscrits extraits de la base NIST. La base d'apprentissage initiale est constituée de 60000 fichiers de chiffres, que nous avons divisé en deux parties :

- une base d'apprentissage qui comporte 40000 chiffres.
- une base de validation qui comporte 20000 chiffres.

La base de test se compose de 10000 chiffres.

5.1 Traitement des données et expérimentation

On extrait les images de la base en éliminant le blanc autour de chiffres et en les normalisant à 24×24 pixels. Un exemple de chiffre après ce traitement est donné dans la figure Fig.5.

Nous avons de plus effectué un redressement des chiffres

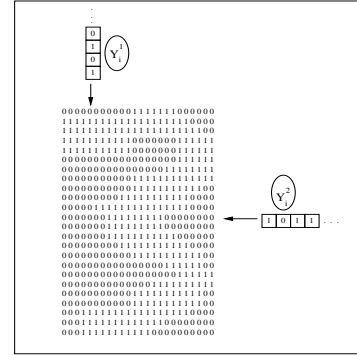


FIG. 5 – Séquences d'observations sur le chiffre 3

suivant l'axe principal d'inertie. Toutes les expériences sont faites à l'aide de la Toolbox Bayesnet écrite sous Matlab [MUR 03]. Ces expériences ont été réalisées dans le cas dynamique avec 10 classes (modèles).

Dans un premier temps nous avons utilisé les HMMs sous forme de RBD pour la recherche du nombre optimal d'états cachés et pour l'évaluation des taux de reconnaissance. Ainsi chaque classe du chiffre pour les modèles HMMs a son propre nombre optimal d'états cachés comme le montre les Tableaux Tab.1 et Tab.2. Ces nombres d'états résultants seront utilisés dans la suite pour les deux cas de couplages.

| | | | | | | | |
|--------------|----|----|----|----|----|----|----|
| classe | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| nbre d'états | 12 | 10 | 13 | 12 | 14 | 14 | 14 |
| classe | 7 | 8 | 9 | | | | |
| nbre d'états | 12 | 13 | 11 | | | | |

TAB. 1 – Nombre d'états optimal par classe pour le HMM-vertical

| | | | | | | | |
|--------------|----|----|----|----|----|----|----|
| classe | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| nbre d'états | 12 | 9 | 14 | 15 | 14 | 15 | 13 |
| classe | 7 | 8 | 9 | | | | |
| nbre d'états | 12 | 14 | 12 | | | | |

TAB. 2 – Nombre d'états optimal par classe pour le HMM-horizontal

Nous signalons que toutes les configurations des états cachés sont autorisées (type ergodique dans le cas des HMMs), pour tous les modèles.

5.2 Résultats

Après avoir effectué les pré-traitements, l'apprentissage des paramètres (construction des modèles), nous passons à la

| | Taux de reconnaissance |
|-----------------|------------------------|
| HMM-vertical | 92.85 % |
| HMM-horizontale | 90.17 % |
| coup1 | 93.79 % |
| coup2 | 94.23% |

TAB. 3 – Résultats de reconnaissance pour les différents modèles

phase de reconnaissance des chiffres manuscrits. Nous considérons 10 modèles (1 par classe). La classe reconnue est celle qui a la vraisemblance $P(O | classe)$ maximum, où O est l'ensemble des séquences associées à l'image du chiffre test. Les résultats concernant les taux de reconnaissance sont donnés dans le tableau Tab.3.

Nous avons utilisé une seule gaussienne pour la modélisation des observations car nous avons remarqué que l'utilisation de plusieurs gaussiennes conduit à des résultats similaires. La seule différence réside dans le nombre des paramètres à calculer, ie., le temps de calcul.

6 Conclusion

Dans ce travail, nous comparons expérimentalement les deux modèles de couplage : coup1 et coup2 avec les HMMs vertical et horizontal. Tous les modèles (HMMs simples et modèles de couplage) ont été réalisés sous forme de réseaux bayésiens dynamiques. Ainsi nous déduisons que les deux modèles de couplage sont plus performants que les modèles basés sur les HMMs simples (voir Tab.3). Cela est dû au fait que les HMMs simples prennent en considération uniquement l'information suivant les lignes ou bien suivant les colonnes et que les modèles couplés regroupent ces deux informations. De plus les modèles de couplage donnent plus d'importance aux observations colonnes qu'aux observations lignes (nature de l'écrit). Ce qui est traduit par les directions des flèches des variables d'états colonnes vers les variables d'états lignes. Le modèle coup2 accorde encore plus d'importance aux observations colonnes qu'aux observations lignes, ce qui est traduit par un meilleur taux de reconnaissance que le modèle coup1. Ainsi ces résultats obtenus sont très encourageants. Nous remarquons que ces modèles de couplage ont aussi été appliqués à la reconnaissance de caractères imprimés [HAL 04]. Les temps d'apprentissage et de reconnaissance sont cependant très longs dans l'implémentation actuelle, comparés à ceux des HMMs classiques avec la boîte à outils HTK. De plus ces résultats peuvent être améliorés en résolvant quelques problèmes liés aux aspects logiciels. D'autres détails demandent encore à être améliorés, surtout en ce qui concerne les conditions initiales, les algorithmes d'inférence et la structure optimale. Pour ce dernier point des travaux sont en cours pour trouver d'autres stratégies structurelles (graphes) qui pourraient améliorer les performances obtenues.

Références

[CHO 03] CHO S., KIM J. H., Bayesian Network Modeling of Hangul Characters for on-Line Handwriting Recognition, *Proceeding of the Seventh International*

Conference on Document Analysis and Recognition (IC-DAR'03), 2003.

[DAO 01] DAOUDI K., FOHR D., ANTOINE C., Continuous Multiband Speech Recognition using Bayesian Networks, *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2001.

[DEV 01] DEVIREN M., DAOUDI K., Structural learning of dynamic Bayesian Networks in Speech Recognition, *EUROSPEECH*, 2001.

[GEO 04] GEOFFROIS E., CHEVALIER S., PRÊTEUX F., Programmation dynamique 2D pour la reconnaissance de caractères manuscrits par champs de Markov, *Reconnaissance des Formes et Intelligence Artificielle*, janvier 2004.

[HAL 02] HALLOULI K., LIKFORMAN L., SIGELLE M., A comparative study between decision fusion and fusion data in Markovian printed character recognition, *Proceedings of 16th International Conference on Pattern Recognition*, 2002, pp. 147-150.

[HAL 03] HALLOULI K., LIKFORMAN-SULEM L., SIGELLE M., Réseaux Bayésiens Dynamiques pour la reconnaissance des caractères imprimés dégradés, *GRETSI, CD-ROM*, 2003.

[HAL 04] HALLOULI K., Reconnaissance de caractères par méthodes markoviennes et réseaux bayésiens, PhD thesis, École Nationale supérieure des Télécommunications, Mai, 2004.

[JEN 88] JENSEN F. V., LAURITZEN S. L., OLESEN K., Bayesian updating in causal probabilistic networks by local computations, *Comp. Stat. Quart.*, vol. 4, 1988, pp. 269-282.

[JEN 94] JENSEN F., JENSEN V., Optimal junction trees, *Proceeding of the 10th conference on Uncertainty in Artificial Intelligence*, 1994, pp. 360-366.

[LAU 99] LAURITZEN L. S., SPIEGELHALTER J., COWELL G. R., DAWID P. A., *Probabilistic Networks and Expert Systems*, Springer, première édition, 1999.

[MUR 02] MURPHY K., Dynamic Bayesian Networks : Representation, Inference and Learning, PhD thesis, University of California, Berkeley, 2002.

[MUR 03] MURPHY K., Bayes Net Toolbox for Matlab, Available on the web at <http://www.ai.mit.edu/murphyk/Bayes/bnintro.html>, October, 2003.

[PAR 98] PARK H., LEE S., A Truly 2-D Hidden Markov Model for Off-Line Handwritten Character Recognition, *Pattern Recognition*, vol. 31, 1998, pp. 1849-1864.

[SAO 97] SAON G., Modèles Markoviens uni et bidimensionnels pour la reconnaissance de l'écriture manuscrite hors-ligne, PhD thesis, Université Henri Poincaré - Nancy1, 1997.

[SOU 02] SOUAFI S., Contribution à la reconnaissance des structures des documents écrits : Approche probabiliste, PhD thesis, INSA de LYON, 2002.

[ZWE 98] ZWEIG G., Speech Recognition with Dynamic Bayesian networks, PhD thesis, University of California, Berkeley, 1998.