

Apprentissage croisé en reconnaissance analytique de l'écriture manuscrite

Christophe Choisy, Abdel Belaïd

► **To cite this version:**

Christophe Choisy, Abdel Belaïd. Apprentissage croisé en reconnaissance analytique de l'écriture manuscrite. Colloque International Francophone sur l'Écrit et le Document - CIFEd'00, 2000, Lyon, France, Presses polytechniques et universitaires romandes, 8 p, 2000, Collection des sciences appliquées de l'INSA de Lyon. <inria-00099039>

HAL Id: inria-00099039

<https://hal.inria.fr/inria-00099039>

Submitted on 26 Sep 2006

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.

APPRENTISSAGE CROISÉ EN RECONNAISSANCE ANALYTIQUE DE L'ÉCRITURE MANUSCRITE

Christophe Choisy et Abdel Belaïd

LORIA/CNRS, Campus scientifique

BP 239, 54506 Vandoeuvre-les-Nancy cedex, France

RÉSUMÉ

Ce papier présente une méthode d'apprentissage croisé de lettres dans le cadre d'une approche analytique de reconnaissance de mots. Les mots sont représentés par des HMMS où chaque état représente une lettre. Chaque lettre est décrite par un champ de Markov causal. La génération des modèles de mots est obtenue par fusion des modèles de lettres. La réestimation des paramètres de ces modèles, objet de cet article, est réalisée au travers de l'apprentissage des modèles de lettres et des transitions des HMMS représentant les mots. L'apprentissage des lettres se fait par croisement des paramètres extraits des différents modèles des mots les contenant, de manière à exploiter tous les contextes d'écriture possibles de ces lettres. L'utilisation de l'algorithme de *Baum-Welch* permet une réestimation optimale de ces paramètres tout en supprimant la nécessité d'une segmentation. Les transitions des HMMS représentant les mots sont également réestimées à l'aide des informations des modèles générés. Les premiers tests, effectués sur 7031 mots de montants de chèques fournis par la SRTP (2/3 en apprentissage, 1/3 en reconnaissance), donnent un taux de reconnaissance de 83.4%.

Mots-clés: HMM, NSHP-HMM, Apprentissage croisé, Algorithme de *Baum-Welch*, Méta-modèles.

1. INTRODUCTION

Les travaux récents en reconnaissance stochastique de l'écriture s'orientent de plus en plus vers des approches 2D; celles-ci prennent en effet mieux en compte la nature fondamentalement 2D de l'écriture [1,2,3,4,5]. Les estimateurs sont généralement des réseaux de neurones (RN), ou des HMMS pseudo-2D (PHMM).

Les RN sont généralement appliqués soit sur des lettres [5] soit sur des segments [3]. L'inconvénient majeur des RN est leur manque d'élasticité, ne permettant pas un recalage des informations en entrée. Les PHMMs, appliqués avec succès dans de nombreux travaux [1,2], s'affranchissent de cette limite : basés sur

un HMM principal selon un axe de l'écriture (par exemple l'horizontale) et des HMMs secondaires selon l'autre axe (les colonnes), ils ont des propriétés d'élasticité 2D. Cependant, ils nécessitent une hypothèse d'indépendance entre colonnes qui n'est pas toujours vraie en pratique.

G. SAON a proposé dans [6] un nouveau modèle basé sur des champs de Markov : le NSHP-HMM. Son architecture, utilisant un HMM, lui confère une élasticité horizontale lui permettant de s'adapter à la longueur des échantillons analysés. Ce modèle s'applique directement sur des images binaires, qu'il analyse colonne par colonne; l'analyse d'un point prend en compte un voisinage 2D de pixels, lui permettant de se libérer de l'hypothèse d'indépendance des colonnes des PHMMs. Il a montré de très bonnes performances en reconnaissance de mots de montants de chèques [7,8].

Ce modèle a été appliqué selon une approche globale des mots de type modèle discriminant. Cette approche a ses limites : l'une est la quantité de paramètres du NSHP-HMM, nécessitant d'importantes bases d'apprentissage (cf chap. 2). Les autres, inhérentes à l'approche globale, sont la limitation à un vocabulaire réduit et distinct. Pour dépasser ces limites, une approche analytique est proposée : les modèles de lettres étant plus petits que ceux des mots, la quantité de paramètres se trouve nettement diminuée; de plus, cette approche permet d'étendre le vocabulaire a priori sans limitations.

De nombreux travaux en reconnaissance analytique de l'écriture se basent sur une segmentation en graphèmes [9,3,10]; cette segmentation, généralement basée sur des critères topologiques, ne peut être totalement fiable [9]. De fait, il semble plus intéressant de laisser soin au système de décider lui-même des meilleures limites pour les lettres.

Alors que la plupart des travaux s'appuient sur des algorithmes de type programmation dynamique pour l'apprentissage et la reconnaissance, notre système utilise l'algorithme de *Baum-Welch*, qui garantit un optimum local.

La méthode proposée est une génération dynamique de modèles de mots, basée sur les modèles de lettres et des HMMs représentant l'enchaînement de ces lettres. La réestimation des modèles de lettres et des transitions entre les lettres se fait par apprentissage croisé. Cette technique, dérivant directement des formules de *Baum-Welch* et déjà utilisée en reconnaissance de la parole, consiste à croiser les informations relatives aux lettres dans les différents modèles de mots. L'utilisation de l'algorithme de *Baum-Welch* permet au système de déterminer la meilleure répartition des paramètres dans les modèles, à partir de la simple connaissance des labels des classes de mots servant à l'apprentissage.

2. LE MODÈLE NSHP-HMM

Le NSHP-HMM (NON-SYMMETRIC HALF-PLANE HIDDEN MARKOV MODEL) est un modèle stochastique de type champ de Markov, architecturé autour d'un HMM. Les états du HMM (états du NSHP-HMM) observent les différentes colonnes d'une image, la probabilité d'observation étant le produit des probabilités de chaque pixel de la colonne. La probabilité d'un pixel est conditionnée par un voisinage de points pris dans le demi-plan non symétrique analysé avant ce pixel, ce qui permet de solutionner le problème de corrélation entre colonnes adjacentes. La figure 1 montre un exemple d'application d'un tel modèle sur une image de lettre. Les méthodes d'apprentissage et de reconnaissance des NSHP-HMM sont décrites dans [6].

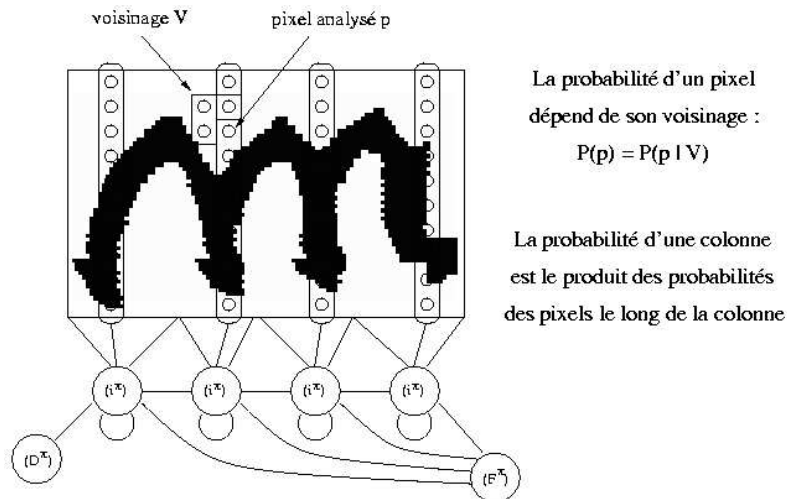


Figure 1 Exemple de NSHP-HMM appliqué sur une lettre, associé à un méta-état x

Les paramètres déterminants du NSHP-HMM sont la hauteur des colonnes analysées (hauteur du modèle), la taille du voisinage (ordre du modèle) et le nombre d'états du HMM sous-jacent. Pour améliorer la répartition des observations (les colonnes) dans les états, deux états tampons D et F ont été ajoutés aux HMMs utilisés; ils permettent de modéliser la probabilité de commencer et de terminer en chaque état.

3. MODÉLISATION DES MOTS

Pour modéliser les mots, nous utilisons des méta-HMMs dans lesquels chaque méta-état représente une lettre (cf figure 2). Pour des raisons techniques, il n'est pas possible de boucler directement sur un état.

Soit m un méta-modèle représentant un mot, ayant S_m états normaux et D_m et F_m comme états tampon. à chaque méta-état $x \in S_m$ est associée une lettre $l(x)$ à laquelle correspond un NSHP-HMM l ayant S^{x^m} états normaux et les états tampon D^x et F^x . On note i^x l'état i du modèle de lettre associé au méta-état x , et D^x et F^x les états tampons de ce modèle (cf figure 1).

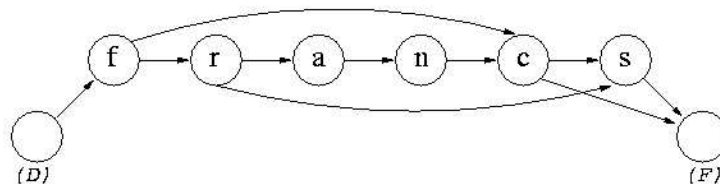


Figure 2 Exemple d'architecture du méta-modèle pour le mot *francs* intégrant des erreurs fréquentes (absence du "s") et des raccourcis d'écriture ("frs", "fcs")

Partant d'un méta-modèle, un NSHP-HMM global est construit en connectant les NSHP-HMMs associés aux lettres de chaque méta-état. Chaque séquence d'états de type $i^x \rightarrow F^x \rightarrow D^y \rightarrow j^y$ est remplacée par une transition $i^x \rightarrow j^y$, dont la valeur est le produit des transitions entre les états de la séquence :

$$P(j^y | i^x) = P(j^y | D^y) * P(D^y | F^x) * P(F^x | i^x)$$

Suivant le même principe, nous avons :

$$P(i^x | D_m) = P(i^x | D^x) * P(D^x | D_m)$$

$$P(F_m | i^x) = P(F_m | F^x) * P(F^x | i^x)$$

La figure 3 montre les étapes de création d'un NSHP-HMM de mot, par remplacement des méta-états par les NSHP-HMMs de lettre correspondant et transformation des transitions (seuls les états des NSHP-HMMs sont représentés). Le modèle ainsi obtenu est un NSHP-HMM qui peut être appliqué comme un modèle global de mot. La limitation interdisant le bouclage sur un méta-état s'explique par le fait que cela générerait des transitions entre les états du NSHP-HMM associé, qui écraseraient alors celles déjà existantes.

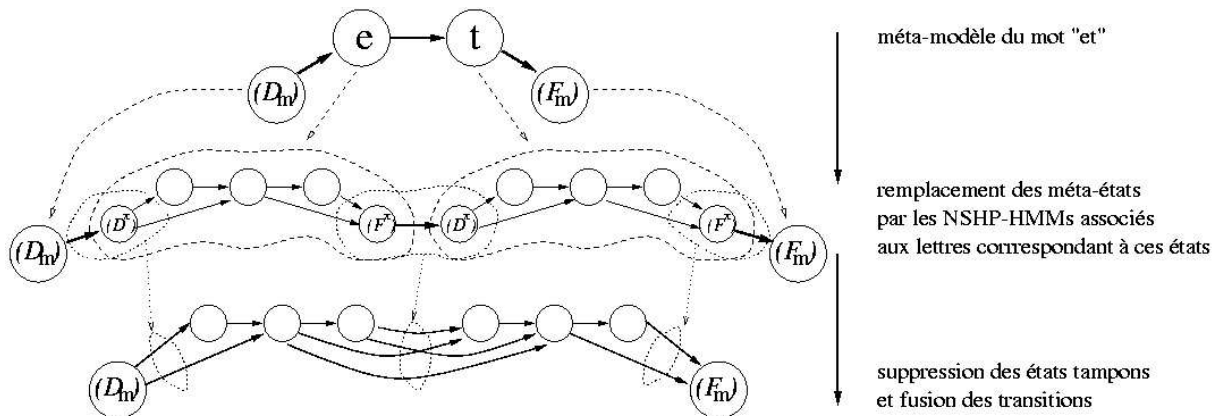


Figure 3 Principe de construction du NSHP-HMM global correspondant au mot "et" à partir du méta-modèle du mot "et" et des NSHP-HMM associés au "e" et au "t"

4. APPRENTISSAGE CROISÉ

4.1 Réestimation des modèles de lettres

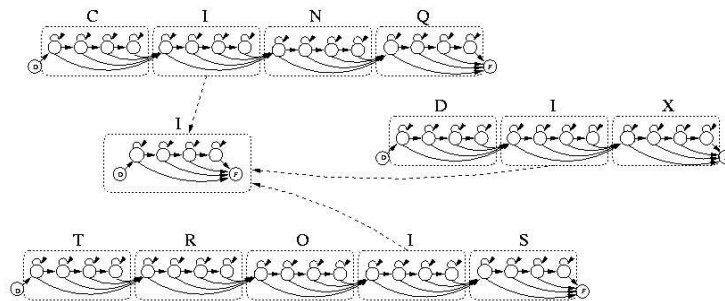


Figure 4 Exemple d'apprentissage croisé du modèle de "i"

L'apprentissage croisé consiste à réestimer les modèles de lettres en utilisant les résultats de la réestimation des modèles de mots. Cette méthode dérive directement de l'apprentissage *Baum-Welch*, en considérant que chaque état de chaque modèle de mot appartient également à un modèle de lettre. La figure 4 illustre ce

principe. La réestimation des transitions entre les états tampon et les états normaux des modèles de lettres se fait au travers des transitions générées entre les lettres.

Pendant la construction d'un modèle global de mot, les états tampon D^x et F^x des modèles de lettres disparaissent; la réestimation des transitions $a_{D^x i^x}$ et $a_{i^x F^x}$ se fait alors au travers des liaisons générées. Soit un méta-modèle m ; K est le nombre d'échantillons (images) analysés, O^k le k ième échantillon, T_k le nombre de colonnes de l'échantillon k ; $P_k = P(O^k | m)$.

Pour deux méta-états $x \neq y$ reliés dans un méta-modèle :

- la transition $a_{D^x i^x}$ est intégrée dans les transitions $a_{j^y i^x}, y \neq x$ et $a_{D_m i^x}$
- la transition $a_{i^x F^x}$ est intégrée dans les transitions $a_{i^x j^y}, y \neq x$ et $a_{i^x F_m}$
- les transitions internes au modèle de lettre $a_{i^x j^x}$ restent inchangées.

Le principe de la réestimation croisée est de recueillir ces informations pour tous les modèles associés à une même lettre dans les différents méta-modèles. Connaissant, par construction, la structure des modèles globaux générés, il est possible de retrouver pour un état d'un modèle généré le modèle de lettre dont il provient, et l'état correspondant dans ce modèle. Il suffit alors d'appliquer l'algorithme de *Baum-Welch* sur tous les modèles générés et de rassembler les résultats concernant cet état pour le réestimer.

En ce qui concerne les transitions entres états des modèles de lettres, deux cas se distinguent : les transitions entre états normaux et les transitions vers les états tampons : $a_{D^x i^x}$ et $a_{i^x F^x}$. Les premières se réestiment de manière analogue aux états. Les secondes nécessitent une sommation de tous les chemins contenant une transition construite grâce à elles. Ainsi, pour la réestimation de $a_{D^x i^x}$:

Soit $W_{D^x i^x}^m$ la somme pondérée sur tous les chemins utilisant cette transition dans le modèle m :

$$W_{D^x i^x}^m = \sum_{k=0}^K \frac{1}{P_k} \left[\sum_{t=0}^{T_k-1} \sum_{y=1}^{S_m} \sum_{\substack{j^y=1 \\ y \neq x}}^{S^y} \alpha_t(j^y) a_{j^y i^x} b_{i^x}(O_{t+1}^k) \beta_{t+1}(i^x) + a_{D_m i^x} b_{i^x}(O_0^k) \beta_0(i^x) \right]$$

Suivant les formules de *Baum-Welch*, les transitions depuis l'état tampon D vers un état normal i pour un modèle NSHP-HMM correspondant à la lettre l sont réestimées comme suit :

$$\overline{a_{Di}^l} = \frac{\sum_{m=1}^M \sum_{x \in S^m} W_{D^x i^x}^m}{\sum_{m=1}^M \sum_{x \in S^m} \sum_{\substack{j^x=1 \\ l(x)=l}}^{S_m} W_{D^x i^x}^m}$$

Intuitivement, cela revient à faire le rapport de tous les chemins passant par la transition $a_{D^x i^x}$ par la somme des chemins passant par l'état D^x , qui est en fait la somme de tous les chemins passant par les transitions $a_{D^x j^x}$ pour tout j^x , et ce pour tous les modèles de mots contenant la lettre l .

4.2 Réestimation des méta-modèles

La génération des modèles globaux de mots se fait à partir de méta-modèles. étant des HMM, on peut réestimer les transitions entre les méta-états à partir des informations des modèles de mots générés. En effet, pour deux méta-états

$x \neq y \in S$, nous avons par construction $a_{xy} = a_{F^x D^y}$, $a_{D_m x} = a_{D_m D^x}$, $a_{x F_m} = a_{F^x F_m}$ (cf figure 3).

- la transition a_{xy} est utilisée pour construire les transitions $a_{i^x j^y}$
- la transition $a_{D_m x}$ est utilisée pour construire les transitions $a_{D_m i^x}$
- la transition $a_{x F_m}$ est utilisée pour construire les transitions $a_{i^x F_m}$

Suivant le même principe que pour la réestimation des modèles de lettres, l'application des formules de *Baum-Welch* permet de réestimer chaque méta-transition en sommant les informations sur tous les chemins l'utilisant.

5. EXPÉRIMENTATIONS

Le système a été testé sur une base de 7031 mots issus de montants de chèques français, fournis par la SRTP¹. Les paramètres des NSHP-HMM des modèles de lettres sont : hauteur de 20 pixels, voisinage de 3 pixels; le nombre d'états normaux pour un NSHP-HMM correspondant à une lettre est $\bar{n}/2+1$, où \bar{n} est le nombre moyen de colonnes pour les échantillons de lettre.

Pour chaque méta-modèle, quatre NSHP-HMM sont créés, correspondant aux quatre symétries possibles des images. La probabilité d'une image est le produit des probabilités obtenues par chaque modèle. Les méta-modèles synthétisent les erreurs fréquentes des mots (cf figure 2).

Deux prétraitements sont appliqués : une correction de la pente de l'écriture, comme proposée dans [8], et une normalisation différentielle, ramenant l'image à 20 pixels de hauteur en normalisant les trois bandes de l'écriture dans trois parties verticales égales.

Deux tests ont été effectués. Le premier valide le principe de l'apprentissage croisé : l'intérêt de cette méthode est que tous les modèles et méta-modèles peuvent théoriquement être appris en même temps.

apprentissage croisé	top 1	top 2	top 3	top 5
5 passes	80.96%	88.48%	91.43%	94.47%
10 passes	83.12%	89.23%	92.35%	95.14%
15 passes	83.41%	89.31%	92.02%	94.84%
20 passes	82.83%	89.15%	91.56%	94.43%
30 passes	82.58%	88.90%	91.64%	94.43%
40 passes	82.20%	88.52%	91.14%	93.97%
50 passes	82.33%	88.36%	91.06%	93.60%

Tableau 5 Taux de reconnaissance moyen pour différents nombres de passes d'apprentissage

¹Service de Recherche Technique de la Poste

La base est décomposée en $2/3$ (4626 mots) pour l'apprentissage croisé et $1/3$ (2405 mots) pour les tests de reconnaissance. Les méta-modèles et les modèles de lettres sont initialisés avec des probabilités d'observation et de transition égales, et plusieurs passes d'apprentissage croisé sont effectuées. Les résultats sont indiqués dans le tableau 5.

Les résultats sont relativement bons, si l'on considère qu'aucune connaissance a priori n'a été fournie au système. On remarque une montée puis une décroissance des probabilités; cela est dû à un phénomène de surapprentissage, détectable au fait que la probabilité des échantillons d'apprentissage continue de croître.

Un second test a été effectué avec préapprentissage des modèles de lettres. La base est séparée en deux ensembles : le premier contient 759 mots segmentés manuellement en 3896 lettres; le second contient les 6272 mots restants.

L'ensemble d'apprentissage des lettres est relativement petit, et le nombre d'échantillons très variable selon les lettres (539 échantillons pour 'n', et seulement 12 pour 'p'); de fait, l'initialisation est pauvre. L'initialisation consiste en 5 passes d'apprentissage des modèles de lettres sur les échantillons de cet ensemble.

Les 6272 mots du second ensemble sont répartis en un ensemble d'apprentissage de 3867 mots et 2405 mots pour les tests, divisant la base initiale dans des conditions similaires au premier test ($2/3 - 1/3$). Les résultats pour diverses passes d'apprentissage croisé sont reportés dans le tableau 6.

apprentissage croisé	top 1	top 2	top 3	top 5
sans (lettres seulement)	72.78%	80.75%	84.55%	89.14%
5 passes	79.92%	87.35%	90.90%	93.86%
10 passes	81.50%	88.10%	90.48%	93.82%
15 passes	81.00%	87.56%	90.02%	93.11%
20 passes	80.54%	87.52%	89.81%	92.69%
30 passes	80.13%	86.47%	89.14%	91.90%
40 passes	80.13%	86.43%	88.77%	91.52%
50 passes	79.67%	85.93%	88.52%	91.11%

Tableau 6 Taux de reconnaissance moyen pour différents nombres de passes d'apprentissage

On constate que les résultats sont inférieurs, ce qui peut paraître surprenant. Cependant, sur la base d'apprentissage les résultats s'avèrent comparables pour les deux tests effectués. Ces résultats inférieurs sont vraisemblablement dûs à deux facteurs : d'une part la base d'apprentissage des lettres est de taille relativement faible, ce qui peut fausser la représentativité des modèles ainsi appris et également entraîner un surapprentissage initial.

D'autre part, cette base étant retranchée de la base d'apprentissage, la taille de cette dernière est considérablement diminuée (16.4% de moins), ce qui augmente le phénomène de surapprentissage. Ce dernier point se remarque dans le décalage à 10 passes d'apprentissage des meilleurs résultats, pour 15 passes lors du premier test (les 5 passes d'apprentissage des modèles de lettres ayant vraisemblablement déjà contribué à ce surapprentissage). D'autres tests devront être faits pour confirmer ces hypothèses.

L'approche analytique permet de réduire la complexité du système proposé par G. SAON [6,8] : le nombre de calculs flottants est en effet proportionnel au nombre d'états des modèles, chacun analysant chaque colonne de l'image. L'approche globale proposée par G. SAON considère la taille moyenne des mots, utilisant ainsi

beaucoup d'états; considérant la moyenne des lettres, notre approche permet de diviser ce nombre d'états par un facteur 7.

Dans un même temps, nous avons observé qu'un voisinage de 4 points était trop important pour modéliser correctement les lettres. 3 points donnent de meilleurs résultats, divisant ainsi par 2 le nombre de paramètres à réestimer, ce qui donne au total une réduction d'un facteur 14 des paramètres à apprendre.

6. CONCLUSION

Nous avons proposé une nouvelle approche pour la reconnaissance analytique de mots, basée sur la génération dynamique de modèles globaux. Cette approche permet une réduction de complexité d'un facteur 14 par rapport au système proposé par G. SAON. Les premiers tests donnent des résultats très encourageants. L'apprentissage des modèles de lettres se fait au travers des modèles de mots, et l'utilisation de l'algorithme de *Baum-Welch* assure un apprentissage optimal. Des tests plus complets doivent être effectués sur des bases plus importantes afin de valider ce système.

Les modèles de mots sont générés dynamiquement d'après les méta-modèles. On remarque que cette méthode peut aisément être étendue à des montants complets grâce à un niveau supplémentaire de méta-modèles, ce qui permettrait d'éliminer la segmentation en mots. Cette extension nécessite cependant de pouvoir trouver la meilleure séquence de mots, si l'on veut éviter de générer tous les montants possibles. Ce problème est du même ordre qu'une généralisation du système à un vocabulaire illimité : il faut trouver la meilleure séquence d'états dans le méta-modèle qui décrit le mot analysé. Ce point nécessite des recherches afin de ne pas perdre les avantages de l'algorithme de *Baum-Welch*.

BIBLIOGRAPHIE

- [1] O. E. Agazzi and S. Kuo, "Hidden Markov Model Based Optical Character Recognition in the Presence of Deterministic Transformation", *Pattern Recognition*, p1813–1826, Feb. 1993.
- [2] R. Bippus, "1-Dimensional and Pseudo 2i-Dimensional HMMs for the Recognition of German Literal Amounts.", *ICDAR'97*, vol 2 p487–490, Aug. 1997.
- [3] B. Lemarié, M. Gilloux, and M. Leroux, "Un modèle neuro-markovien contextuel pour la reconnaissance de l'écriture manuscrite", *actes 10ème Congrès AFCET de RFIA*, 1996.
- [4] H.S. Park and S.W. Lee, "SAN HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition", *IEEE Proceedings of ICPR'96*, vol 2 p320–324, 1996.
- [5] J.C. Simon, O. Barret, and N. Gorski, "A System for the Recognition of Handwritten Literal amounts of checks", *DAS'94*, p135–155, Sept.1994.
- [6] G. Saon, "Modèles markoviens uni- et bidimensionnels pour la reconnaissance de l'écriture manuscrite hors-ligne", *PhD thesis*, 1997.
- [7] G. Saon and A. Belaïd, "High Performance Unconstrained Word Recognition System Combining HMMs and Markov Random Fields", *IJPRAI*, p771–788, 1997.
- [8] G. Saon and A. Belaïd, "OffP-line Handwritten Word Recognition Using a Mixed HMM-MRF Approach", *ICDAR'97*, vol 1 p118–122, 1997.
- [9] M-Y. Chen, A. Kundu, and J. Zhou, "Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type stochastic Network", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, p481–497, 1994.
- [10] F. Kimura, M. Shridhar, G. Houle, "Handwritten word recognition using lexicon free and lexicon directed word recognition algorithms", *ICDAR'97*, 1997.