

Couplage d'une vision locale par hmm et globale par RN pour la reconnaissance de mots manuscrits

Christophe Choisy, Abdel Belaïd

► **To cite this version:**

Christophe Choisy, Abdel Belaïd. Couplage d'une vision locale par hmm et globale par RN pour la reconnaissance de mots manuscrits. Colloque International Francophone sur l'Écrit et le Document - CIFED'2002, 2002, Hammamet, Tunisie, 10 p, 2002. <inria-00099447>

HAL Id: inria-00099447

<https://hal.inria.fr/inria-00099447>

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.

Couplage d'une vision locale par HMM et globale par RN pour la reconnaissance de mots manuscrits

Ch. Choisy — A. Belaid

LORIA

Campus scientifique, BP 239
54506 Vandoeuvre-lès-Nancy Cedex, France
choisy,abelaid@loria.fr

RÉSUMÉ. Nous présentons dans cet article une idée de combinaison de modèles à visions locale et globale. Séparément, ces deux types de modèles ont prouvé leurs capacités, et leur combinaison a été exploitée en utilisant les modèles globaux plutôt localement, et les modèles locaux pour synthétiser leurs résultats. Nous proposons une démarche inverse en utilisant les modèles locaux pour normaliser efficacement les images de manière non linéaire, et les modèles globaux pour assurer une cohérence sur l'ensemble de la forme. Les modèles locaux sont de type markovien, et les modèles globaux sont des réseaux de neurones. Le gain de la normalisation markovienne est de l'ordre de 3% par rapport à une normalisation linéaire classique (84.47%). Le gain apporté par la combinaison est de l'ordre de 2% par rapport à l'approche markovienne seule (85.35%), soit un score de 87.41%.

ABSTRACT. In this paper, an idea for the combination of global and local view models is presented. These two type of models have proved their capabilities independantly. Some combination were proposed, using global view models for local analysis, and local view models to synthetize local results. An opposite approach is proposed here : local view models are used as a normalization tool, while global view models are used for the recognition of the normalized image. The use of local view models for normalization is justified by their capability to perform a non-linear normalization according to the image information. We propose Markov models as local view models, and Neural Netwok as global view models. Using Markov models for the normalization increases results up to 3% better than a classical linear normalization. Global results are improved, performing 2% better than the Markov model itself.

MOTS-CLÉS : Vision Locale et Globale, Modèles élastiques, Normalisation non-linéaire, HMM, NSHP-HMM, RN.

KEYWORDS: Global and Local Vision, elastic modeling, non-linear modelization, HMM, NSHP-HMM, RN.

1. Introduction

La grande variabilité de l'écriture manuscrite, au niveau des mots particulièrement, explique le succès des modèles élastiques, tels les HMMs, pour sa modélisation et sa reconnaissance, car ils savent bien s'adapter à ces déformations [Bippus 97, Bunke 95, Park 96]. Cette capacité d'adaptation provient d'une vision locale couplée à des méthodes de type programmation dynamique, permettant un recalage sur les observations importantes tout en absorbant les déformations. Cette vision locale fait leur force et leur faiblesse, car elle interdit les corrélations sur l'ensemble de la forme analysée. En effet, Il est supposé que la mesure des dépendances locales entre les observations est représentative de la globalité de l'échantillon analysé. En pratique cette hypothèse n'est pratiquement jamais vérifiée, d'où l'importance de la prise en compte du contexte soulevée dans certains travaux [Lemarié 96].

Pour la reconnaissance de caractères isolés, les modèles globaux, tels les réseaux de neurones, s'avèrent beaucoup plus performants [Gilloux 95, LeCun 98, Vapnik 95]. Cela s'explique par leur vision globale permettant des corrélations sur l'ensemble de la forme. Mais ces modèles sont très sensibles aux distorsions [Alexandre 94], et ne peuvent analyser que des données de taille fixe.

Pour exploiter la force des ces deux types de modèles tout en dépassant leurs limites, nous proposons d'utiliser les modèles à vision locale pour normaliser les images, avant analyse par un modèle à vision globale. Nous utilisons un champ de Markov développé dans l'équipe, le NSHP-HMM [Saon 97], pour normaliser les images de manière non-linéaire, et deux étages de RN pour la reconnaissance globale des images résultantes. Les résultats obtenus confirment largement l'intérêt de ce couplage.

Pour caractériser ces deux types de modèles, nous avons introduit la notion de "Modèles à Vision Locale" (MVL) pour les modèles élastiques, et de "Modèles à Vision Globale" (MVG) pour les modèles de type RN et SVM [Choisy 01]. Ici, nous redéfinissons clairement notre notion de global et de local :

Vision locale : Dans ce type de modèle, les observations sont supposées indépendantes entre elles. L'analyse d'une forme dans son ensemble se fait par maximisation de la probabilité d'estimation de la forme, d'après la probabilité des observations locales et des contraintes locales entre observations. Il n'y a pas de réelles corrélations entre les observations sur l'ensemble de la forme.

Vision globale : Dans ce type de modèle, la forme est observée dans sa globalité, permettant des corrélations sur l'ensemble de la forme.

On voit clairement que les modèles élastiques de type HMM sont des modèles à vision locale, tandis que les SVM et les RN sont des modèles à vision globale. Le cas des TDNN [Seni 96] et des RN récurrents [Senior 98] illustre des tentatives de conciliation des RN et de l'élasticité, mais cela se fait au détriment de la vision, qui devient locale. Nous les rangeons donc également dans la catégorie des MVL.

Cinq sections suivent cette introduction. La section 2 décrit de manière générale le système, pour la phase d'apprentissage et de reconnaissance. Les sections 3 et 4 décrivent les modèles local et global utilisés et leurs particularités. La section 5 décrit brièvement les méthodes d'apprentissage des modèles. La section 6 décrit et discute les résultats obtenus, amenant également des perspectives de travail.

2. Description générale du système

Pour simplifier les explications, nous appellerons par la suite N le nombre de classes de mots dans les bases d'apprentissage et de reconnaissance. La figure 1 illustre le schéma général d'apprentissage de notre système.

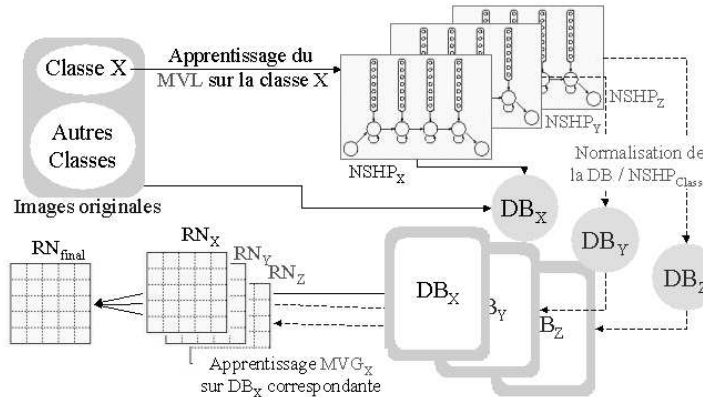


Figure 1: Schéma général du système

Le système se décompose en quatre grandes étapes :

1) **L'apprentissage des modèles à vision locale** : on considère un champ de Markov (NSHP-HMM) par classe de mot, soit N NSHP-HMM. Chacun apprend à modéliser la classe de mot correspondante.

2) **La normalisation** : chaque NSHP-HMM est utilisé pour normaliser toute la base de données. Nous obtenons donc N bases de données (une par classe), chacune contenant les images des N classes normalisées.

3) **L'apprentissage des modèles à vision globale du premier niveau** : on considère un réseau de neurones par banque normalisée, soit N RN de premier niveau. Les réseaux sont entraînés à discriminer les différentes classes dans une base de données normalisée.

4) **L'apprentissage du RN du second niveau** : chaque RN de l'étape précédente produit autant de valeurs que de classes, soit $N * N$ valeurs au total. Nous utilisons donc un nouveau RN pour faire la synthèse de ces résultats.

Le schéma de reconnaissance est très proche du schéma d'apprentissage. Il se décompose en trois étapes :

1) **La normalisation** : l'image à reconnaître est normalisée par chacun des NSHP-HMM. Nous obtenons ainsi N images normalisées.

2) **L'analyse des images normalisées** : les RN du premier niveau analysent chacun l'image normalisée par le NSHP-HMM correspondant. Nous obtenons en sortie du premier niveau N valeurs pour chaque RN, soit $N * N$ valeurs au total.

3) **La synthèse du premier niveau** : le RN du second niveau prend en entrée ces $N * N$ valeurs et fournit N valeurs correspondant à la proximité de l'image d'entrée à chaque classe. Le choix de classification est fait par maximum de vraisemblance.

3. Modèle à vision locale

Nous avons choisi le NSHP-HMM comme modèle à vision locale pour les raisons suivantes :

- Il travaille directement sur l'image, au niveau des colonnes de pixels, permettant une normalisation fine.
- Basé sur un HMM, il sait bien s'adapter aux variations de longueur.
- Sa capacité de modélisation de l'écriture a été mise en évidence au cours de plusieurs travaux, tant en approche globale [Saon 97] qu'analytique [Choisy 00].
- Il ne nécessite pas de caractéristiques fortes, évitant de faire des hypothèses fortes sur des zones de l'image (pas de décision dures). Cela lui offre une grande liberté pour le repérage des indices visuels.

Sur le plan technique, c'est une hybridation entre un HMM et un champ de Markov observable. Le champ estime les probabilités d'observation des colonnes de l'image par les états du HMM ; ce dernier synthétise ces résultats sur l'ensemble de l'image, comblant le manque d'élasticité du champ observable [Saon 97]. Étant donné que le champ n'est pas élastique, le modèle ne fait pas d'adaptation verticale ; la hauteur d'analyse est donc fixe. Pour réduire la variabilité verticale, une normalisation différentielle des images est appliquée : elle consiste à normaliser les trois bandes de l'écriture dans trois zones de taille identique.

Normalisation par NSHP-HMM : La normalisation d'une image par un NSHP-HMM se fait de la manière suivante : l'algorithme de Viterbi permet de récupérer la meilleure répartition des colonnes de l'image dans les états du NSHP-HMM. À chacun de ces états correspond une colonne de l'image normalisée, donnant ainsi une taille fixe à toutes les images. Le contenu de cette colonne est la moyenne des colonnes observées par l'état correspondant dans l'image d'origine [Choisy 01]. La figure 2 illustre ce principe sur un échantillon du mot "et".

La figure 3 illustre l'effet des deux normalisation sur divers échantillons des mots "et" et "cent". La première colonne contient l'image originale, la seconde l'image normalisée par NSHP-HMM, et la troisième l'image normalisée linéairement. On constate une meilleure synchronisation des images dans le cas de la normalisation markovienne. On constate également que la barre des "t" est absorbée par un seul état pour

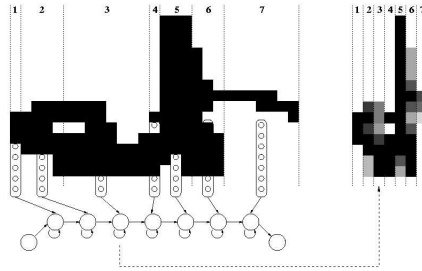


Figure 2: Normalisation du mot “et” par le NSHP-HMM correspondant à cette classe. Le niveau de gris correspond à la densité de points noirs observés à cette position par l’état correspondant à la colonne de l’image normalisée.

la normalisation markovienne, sauf dans le cas des échantillons de “cents”, où le “s” prend également en compte la barre du “t” au dessus de lui.

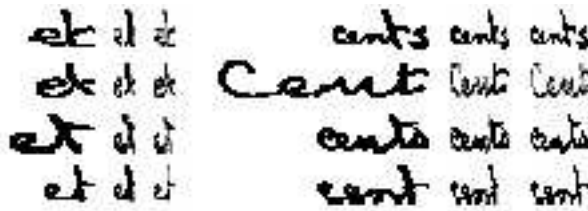


Figure 3: Échantillons de “et” et de “cent” : la première colonne donne l’image originale, la seconde l’image normalisée par NSHP-HMM, la troisième l’image normalisée linéairement.

4. Modèle à vision globale

Nous avons retenu les RN pour la vision globale. Comparativement aux SVM utilisés dans un travail précédent [Choisy 01], ils ont l’avantage de donner un taux d’appartenance aux différentes classes, plutôt qu’une réponse binaire. Le problème du temps d’apprentissage a été résolu en intégrant une topologie spécifique pour les couches cachées, qui permet de réduire considérablement le nombre de liens entre couches. Des tests sur la base NIST de caractères manuscrits ont mis en évidence une amélioration des résultats avec un gain de temps considérable en apprentissage et reconnaissance.

Deux niveaux de RN sont utilisés, comme indiqué en §.2. Le premier niveau se charge de l’apprentissage des images normalisées pour chaque classe. Le second niveau synthétise les résultats du premier et renvoie le score final, utilisé pour la classification.

Les RN du premier niveau : ils possèdent deux couches cachées. La première contient autant de neurones que la couche d'entrée ; la seconde en possède 25 fois moins (valeur empirique, la taille de l'image d'entrée est divisée par 5 horizontalement et verticalement). Les deux couches ont une topologie particulière : chaque neurone observe un carré sur la couche précédente, donnant ainsi une notion 2D à la modélisation des images. La première couche cachée observe des carrés de $5 * 5$ pixels centrés sur chaque neurone de la couche d'entrée. La seconde couche observe des rectangles centrés homogènement sur la couche précédente ; leur taille s'étend jusqu'au milieu des deux centre voisins. La couche finale, estimant l'appartenance aux classes, est complètement connectée à la deuxième couche cachée ; elle contient autant de neurones qu'il y a de classes. Le nombre de neurones sur la couche d'entrée est déterminé par la dimension des images à analyser ; la hauteur de l'image normalisée est donnée par la taille des colonnes analysées par le NSHP-HMM, tandis que la longueur est fixée par le nombre d'états du NSHP-HMM ayant normalisé (§.3) ;

Un RN du premier niveau apprend à discriminer les images des différentes classes normalisées par le NSHP-HMM correspondant à une classe donnée. Il fournit un score de proximité de l'image normalisée à chacune des classes, donc autant de scores que de classes. Lors de la reconnaissance, l'image normalisée par chaque NSHP-HMM est fournie en entrée du RN correspondant ; nous obtenons ainsi pour chaque classe une série de scores, qui sont transmis au niveau suivant.

Le RN du deuxième niveau : c'est un modèle complètement connecté à une couche cachée. Il prend en entrée les résultats des différents RN du niveau précédent, qu'il synthétise pour fournir le score final d'appartenance de l'image à chaque classe.

5. Apprentissage des modèles

Les NSHP-HMM sont entraînés sur la base d'apprentissage selon la technique d'apprentissage croisé décrite en [Choisy 00]. L'idée de base est de synthétiser les informations observées par les modèles de lettres dans les différents modèles de mots. Cela permet une ouverture de notre système sur une approche analytique. Chaque classe de mots est représentée par un NSHP-HMM, qui est entraîné selon un principe de maximum de vraisemblance, grâce à l'algorithme de *Baum-Welch*. Après apprentissage, chaque NSHP-HMM normalise toutes les images de la base d'apprentissage, fournissant ainsi N bases normalisées.

Les RN sont entraînés par une méthode classique de rétropropagation de l'erreur. Chacun des N RN du premier niveau est entraîné sur la base normalisée par le NSHP-HMM correspondant, chaque classe étant représentée par un RN. Une fois l'apprentissage terminé, les RN du premier niveau sont utilisés pour générer une nouvelle base contenant les scores obtenus pour chaque image de la base d'origine, pour chaque RN du premier niveau, soit $N * N$ valeurs pour chaque image. Cette nouvelle base sert de base d'apprentissage au RN du second niveau, chargé de la synthèse des scores.

Une remarque importante est que la base d'apprentissage initiale sert pour tous les apprentissages. Ce point important sera repris dans la discussion des résultats.

6. Résultats

Nous avons appliqué ce concept à la reconnaissance de mots de montants de chèques français. La base de données contient 25249 images de qualité industrielle, réparties dans 28 classes (les 26 mots du vocabulaire des montants plus les abréviations “frcs” et “cts”). Les NSHP-HMM associés aux classes prennent en compte les principales erreurs d’orthographe. 16650 images servent à l’apprentissage des NSHP-HMM et des RN, et 8599 images pour les tests, soit un découpage $2/3 - 1/3$ des classes.

La hauteur d’analyse des NSHP-HMM est fixée à 20 pixels. Pour une classe de mots donnée, le nombre d’états du NSHP-HMM correspond approximativement à la moitié de la longueur moyenne des images de la classe. Cette taille a été retouchée manuellement afin de prendre en compte la topologie des lettres ; une analyse des probabilités de sortie en chaque état a permis de déterminer le nombre d’états optimal pour chaque lettre. La dimension d’entrée du RN correspondant à la classe est déterminée d’après ces informations (§.4) : elle est donc de 20 fois le nombre d’états du NSHP-HMM correspondant. Le RN final a $28 * 28 = 784$ neurones en entrée, et 28 neurones pour la couche cachée et la couche de sortie.

La table 1 permet de comparer les résultats obtenus par le NSHP-HMM lui-même, par notre système pour une normalisation linéaire et pour une normalisation par NSHP-HMM. La normalisation linéaire donne aux images les mêmes dimensions que la normalisation markovienne. Les ensembles d’apprentissage et de test sont les mêmes.

Tableau 1: Scores de reconnaissance pour différentes approches

Normalisation	Reconnaissance	Top 1	Top 2	Top 3	Top 5
sans	NSHP-HMM	85.35%	91.97%	94.39%	96.43%
linéaire	RN	84.47%	90.92%	93.43%	95.50%
NSHP-HMM	RN	87.41%	93.05%	95.41%	96.86%

6.1. Discussion

Les résultats obtenus prouvent la validité de notre approche. En effet, le gain de la normalisation markovienne est de l’ordre de 3% par rapport à une normalisation linéaire (84.47%). Le gain apporté par la combinaison est de l’ordre de 2% par rapport à l’approche markovienne seule (85.35%), soit un score de 87.41%. Les résultats montrent la puissance des modèles à vision globale sur des images correctement normalisées. La normalisation linéaire donnant de moins bons résultats que la normalisation par NSHP-HMM, nous en déduisons l’intérêt de la normalisation par modèles à vision locale.

La table 2 décrit les scores des RN de premier niveau pour les deux normalisations proposées. On constate que la normalisation linéaire obtient la plupart du temps des résultats bien meilleurs à cette étape que la normalisation markovienne. Cependant, la

Tableau 2: Scores par classe des RN du premier niveau : normalisation linéaire (première ligne) et markovienne (seconde ligne)

un	deux	trois	quatre	cinq	six	sept
78.14%	78.75%	79.40%	78.36%	78.75%	79.32%	79.07%
68.00%	73.66%	72.80%	75.52%	73.49%	67.51%	71.08%
huit	neuf	dix	onze	douze	treize	quatorze
77.89%	79.11%	78.35%	79.56%	78.90%	78.90%	77.29%
72.51%	73.14%	70.26%	71.92%	72.82%	74.04%	75.03%
quinze	seize	vingt	trente	quarante	cinquante	soixante
78.17%	79.56%	78.68%	78.61%	77.29%	76.14%	77.99%
75.53%	71.16%	75.08%	77.82%	78.25%	78.14%	77.76%
cent	mille	francs	et	centimes	frcs	cts
78.89%	77.90%	78.36%	77.43%	77.29%	79.07%	79.32%
75.71%	66.23%	74.71%	66.86%	76.59%	69.15%	69.66%

synthèse des résultats donne l'avantage à la normalisation markovienne, ce qui traduit des scores plus complémentaires pour la première couche des RN. Cela suggère que la normalisation markovienne permet une meilleure discrimination inter-classes des images, même si elle pénalise les RN pour la discrimination des classes.

Les résultats obtenus souffrent de limites applicatives. Ainsi, la base de données étant de taille réduite, la même base d'apprentissage a été utilisée à tous les niveaux, depuis le NSHP-HMM pour la normalisation, jusqu'au RN final. La capacité de généralisation des RN s'en trouve réduite. Certaines classes sont sous-représentées, ce qui pénalise fortement les RN ; en effet, une bonne généralisation nécessite l'observation d'un grand nombre d'échantillons. Ce point est à traiter soit à l'aide d'une base plus importante, soit par des méthodes de maximisation de la marge lors de l'apprentissage.

Dans ce travail, les RN apprennent à discriminer à égalité chaque classe. Dans le travail précédemment proposé [Choisy 01], l'utilisation des SVM impliquait une réponse binaire, et les SVM apprenaient à séparer la classe correspondante de toutes les autres. Il serait intéressant d'orienter l'apprentissage des RN dans ce même sens, en apprenant la séparation de la classe courante d'avec les autres, tout en proposant la discrimination des classes. Cela peut se faire par exemple par un renforcement de l'apprentissage de la classe correspondant au RN.

6.2. Perspectives

La normalisation markovienne offre un avantage important sur la normalisation linéaire : on peut l'appliquer au niveau des lettres. Les premiers travaux sont basés sur des mots entiers, mais l'apprentissage croisé permet de retrouver facilement quel état appartient à quel modèle de lettre dans le NSHP-HMM associé à un mot [Choisy 00].

En utilisant ces informations, il est possible d'utiliser les RN sur les lettres normalisées uniquement, et donc d'avoir une approche analytique plutôt que globale. La normalisation linéaire ne permet pas cela, n'ayant pas d'informations sur les limites des lettres.

La méthode de normalisation proposée est assez simple et peut être améliorée. Les niveaux de gris pourraient être affinés en considérant les probabilités des colonnes. Dans le même ordre d'idée, une normalisation basée sur *Baum-Welch* plutôt que *Viterbi* pourrait améliorer cette étape. Les NSHP-HMM peuvent également apporter d'autres informations lors de la normalisation, telles que des probabilités locales, des informations sur l'importance d'une colonne dans l'image, etc.

Dans ce travail, la normalisation est faite au niveau de la colonne de pixels, absorbant ainsi les déformations horizontales. Il serait intéressant d'affiner encore cette normalisation pour pouvoir gérer également les déformations verticales. Un modèle de type PHMM pourrait être employé pour cela. La répartition des états du NSHP-HMM sur les images de sa classe a également besoin d'être améliorée, par exemple en employant des modèles de durée plus fins que les transitions classiques. De manière générale, il sera intéressant d'étudier l'effet du degré de rigidité des MVL sur la normalisation.

Notre approche se décompose en deux parties principales : normalisation et reconnaissance. La première peut être utilisée comme pré-classification lorsqu'on utilise certains types de MVL tels que les modèles markoviens. La deuxième peut venir renforcer cette classification pour les cas difficiles en tirant avantage de la première étape pour la normalisation et pour le niveau de représentation (mot, lettre, etc.).

Une autre voie de recherche intéressante est l'amélioration de l'apprentissage des NSHP-HMM en utilisant la reconnaissance par les RN pour déterminer la meilleure répartition d'états. De manière plus générale, les informations fournies par le MVG (modèle à vision globale) pourraient être utilisées pour améliorer la qualité de la normalisation par les MVL.

7. Conclusion

Ce travail décrit l'utilisation d'un "Modèle à Vision Locale" tel que le NSHP-HMM pour normaliser une image avant reconnaissance par un "Modèle à Vision Globale" tel que les RN. La normalisation proposée s'appuie sur la répartition fournie par l'algorithme de *Viterbi* des états du NSHP-HMM sur l'image analysée.

Malgré une base de donnée limitée, ne permettant pas une bonne généralisation des apprentissages, la supériorité des résultats pour la normalisation markovienne valide le principe proposé.

Le schéma de travail proposé est très général, et de nombreux travaux sont à effectuer pour explorer toutes les possibilités d'une telle approche.

8. Bibliographie

- [Bippus 97] R. Bippus. 1-Dimensional and Pseudo 2-Dimensional HMMs for the Recognition of German Literal Amounts. *ICDAR'97*, vol 2, pages 487–490, Ulm, Germany, Aug. 1997.
- [Bunke 95] H. Bunke, M. Roth et E.G. Schukat-Talamazzini. Off-line Cursive Handwriting Recognition Using Hidden Markov Models. *Pattern Recognition*, 28(9) :1399–1413, 1995.
- [Choisy 00] C. Choisy et A. Belaid. Apprentissage Croisé en Reconnaissance Analytique de l'Écriture Manuscrite. *CIFED'2000*, Lyon, juillet 2000.
- [Choisy 01] C. Choisy et A. Belaid. Handwriting Recognition Using Local Methods for Normalization and Global Methods for Recognition. *ICDAR'2001*, 2001.
- [Alexandre 94] F. Alexandre et F. Guyot. Evaluation d'un modèle connexionniste simple pour la reconnaissance automatique de caractères. *Actes du 3^{ème} Colloque National sur l'Écrit et le Document*, pages 99–108, Rouen, juillet 1994.
- [Gilloux 95] M. Gilloux, B. Lemarié et M. Leroux. A Hybrid Radial Basis Function Network/Hidden Markov Model Handwritten Word Recognition System. *ICDAR'95*, pages 394–397, Montréal, 1995.
- [LeCun 98] Y. LeCun, L. Bottou, Y. Bengio et P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.
- [Lemarié 96] B. Lemarié, M. Gilloux et M. Leroux. Un modèle neuro-markovien contextuel pour la reconnaissance de l'écriture manuscrite. *RFIA'96*, Rennes, France, 1996.
- [Park 96] H. S. Park et S. W. Lee. An HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition. *IEEE Proceedings of ICPR'96*, vol 2, pages 320–324, 1996.
- [Saon 97] G. Saon. *Modèles markoviens uni- et bidimensionnels pour la reconnaissance de l'écriture manuscrite hors-ligne*. Thèse de Doctorat, Université Henri Poincaré - Nancy I, Vandœuvre-lès-Nancy, 1997.
- [Seni 96] Giovanni Seni, Rohini K. Srihari et Nasser M. Nasrabadi. Large vocabulary recognition of on-line handwritten cursive words. *PAMI*, 18(7) :757–762, 1996.
- [Senior 98] A. W. Senior et A. J. Robinson. An Off-Line Cursive Handwriting Recognition System. *PAMI*, 20(3) :308–321, March 1998.
- [Vapnik 95] Vladimir N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.