Normalisation par modèles locaux et reconnaissance par modèles globaux pour la reconnaissance de l'écriture manuscrit

Christophe Choisy, Abdel Belaïd

▶ **To cite this version:**

Christophe Choisy, Abdel Belaïd. Normalisation par modèles locaux et reconnaissance par modèles globaux pour la reconnaissance de l'écriture manuscrit. 13e Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle - RFIA 2002, 2002, Angers, France, 8 p. inria-00099446

HAL Id: inria-00099446
https://inria.hal.science/inria-00099446

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.

Normalisation par modèles locaux et reconnaissance par modèles globaux pour la reconnaissance de l'écriture manuscrite

Normalization by local models and recognition by global models applied to handwriting recognition

Ch. Choisy[1] A. Belaid[1]
[1] LORIA
Campus scientifique, BP 239
54506 Vandoeuvre-lès-Nancy Cedex, France
choisy@loria.fr, abelaid@loria.fr

Résumé

Le principal problème en Reconnaissance de l'Ecriture Manuscrite est la grande variabilité de l'écriture et les distorsions des échantillons. Les modèles élastiques tels que les HMM *sont particulièrement efficaces pour absorber ces variations, grâce à l'utilisation d'observations locales et de programmation dynamique. Mais leur vision de la forme reste locale. D'un autre coté, les modèles globaux tels que les Réseaux de Neurones savent faire des corrélations sur la globalité d'un échantillon. Mais leur entrée de taille fixe ne leur permet pas de s'adapter aux variations de longueur, et ils sont très sensibles aux déformations. Pour utiliser les avantages des deux classes de modèles, nous proposons de normaliser les images à l'aide d'un modèle élastique (un* NSHP-HMM*), puis de les analyser à l'aide d'un modèle global (un* SVM*). Le* NSHP-HMM *se focalise sur les caractéristiques importantes en absorbant les distorsions. L'image est normalisée non-linéairement d'après ces informations, et un* SVM *est utilisé pour les corrélations globales et la classification. Les premiers résultats sont encourageants et tendent à confirmer l'intérêt de notre approche.*

Mots Clef

vision locale et globale, modèles élastiques, normalisation non-linéaire, HMM, SVM.

Abstract

A major problem in handwriting recognition is the huge variability and distortions of patterns. Elastic models such HMM *are efficient to absorb these variations, by use of local observations and dynamic programming. But their vision of the form is local. Furthermore global models such Neural Network are efficient to make correlations on an entire pattern. But their fixed input length cannot face to length variability and they are very sensitive to distortions. To take advantage of these two classes of models, we propose to use an elastic model (a* NSHP-HMM*) to normalize images; a global model (a* SVM*) is used to analyse the resulting samples. The* NSHP-HMM *searchs the important features and absorbs the distortions. The image is normalized non-linearly using these informations. Then a* SVM *is used to estimate global correlations and classify the pattern. The first results are encouraging and tend to confirm the interest of our approach.*

Keywords

local and global view, elastic models, non-linear normalization, HMM, SVM.

1 Introduction

La grande variabilité de l'écriture a rendu l'utilisation des modèles élastiques très populaire ces dernières années. Les plus courants sont les HMM (Hidden Markov Model) [12, 8], les PHMM (Planar HMM) [1] et les champs aléatoires [9, 10], dont l'utilisation en REM (Reconnaissance de l'Écriture Manuscrite) est devenue classique. Leur succès est dû sans conteste à leur capacité d'absorption du bruit et des distorsions, et d'adaptation aux variations de longueur. Cette capacité provient de l'utilisation d'observations locales couplée à des techniques de programmation dynamique, qui permettent à ces modèles de se focaliser sur les caractéristiques importantes de l'écriture. L'hypothèse de base sous-tendant leur utilisation est que la mesure des dépendances locales entre les observations est repréentative de la globalité de l'échantillon analysé. Cette hypothèse n'est pas toujours vérifiée en pratique, mais elle donne des propriétés très utiles à ces modèles. Étant donné le caractère local des observations et des dépendances, nous donnerons à de tels modèles la dénomination de "Modèles à Vision Locale" (**MVL**).

La reconnaissance de mots et de caractères manuscrits fait également appel à d'autres types de modèles tels que les SVM (Support Vector Machine) et les RN (Réseaux de Neurones) [13, 5]. Ces modèles ont une vision globale des échan-

tillons, ce qui leur permet de faire des corrélations entre toutes les parties d'une image. Leurs défauts majeurs sont une grande sensibilité aux déformations importantes [3], ainsi qu'un espace d'entrée de dimension fixe qui ne leur permet pas de s'adapter aux variations de longueur de l'écriture. Étant donné leur vision globale des échantillons, nous appellerons de tels modèles des "Modèles à Vision Globale" (**MVG**).
La plupart des travaux combinant des HMM et des RN [5, 11] utilisent les modèles globaux comme estimateurs locaux du HMM. Nous pensons que les capacités des RN ne peuvent pas donner toute leur mesure dans ces conditions, car les corrélations apprises restent très locales.

L'idée fondamentale de ce travail est de combiner les capacités de ces deux types de modèles, en utilisant un MVL pour la normalisation d'image, et un MVG pour leur reconnaissance. Le MVL permet de rechercher les caractéristiques importantes d'une image tout en absorbant les variations. Connaissant la position de ces caractéristiques, il est possible de normaliser non-linéairement l'image afin de lui donner une taille fixe tout en conservant sa représentativité. L'image normalisée peut alors être analysée par un MVG, qui fera les corrélations entre les différentes parties de l'échantillon.
L'idée intuitive sous-jacente est qu'un modèle à vision locale va rechercher des indices dans une image sans se préoccuper de la cohérence globale de l'image par rapport à ces indices. Ces informations sont alors repositionnées d'après les déformations du MVL, et le modèle à vision globale vient vérifier la cohérence globale des indices relevés. Un MVL non adapté à la forme à reconnaitre ira ainsi relever des indices localement cohérents, mais dont la cohérence globale sera faible.

Nous avons choisi le NSHP-HMM (Non-Symmetric Half Plane HMM) [10] comme MVL. Ce modèle est un HMM dans lequel les probabilités d'observation des états sont estimées par un champ de Markov. Le NSHP-HMM travaille directement sur les images, qu'il analyse colonne par colonne. Cela permet une normalisation fine des images, au niveau de la colonne de pixels. Le HMM confère à ce modèle une propriété d'élasticité lui permettant de s'adapter aux variations de longueur des mots. L'algorithme de Viterbi permet d'obtenir la meilleure répartition des états du HMM sur les colonnes de l'image.
Comme MVG, nous avons retenu le SVM [7]. Les SVM sont basés sur un kernel d'échantillons, qui permet de calculer le meilleur hyperplan séparateur entre 2 classes. Ces modèles sont simples d'utilisation; l'apprentissage et la classification se font rapidement. Appliqués à la reconnaissance de chiffres manuscrits [13] et à la classification de majuscules manuscrites [4], ils ont obtenus des résultats comparables à ceux des RN. Les SVM utilisés permettent uniquement de séparer deux classes.

Bien entendu, les modèles retenus comme MVL et MVG peuvent être remplacés par d'autres. L'objectif de ce travail est avant tout d'explorer le concept de normalisation par un MVL pour une reconnaissance par un MVG. La figure 1 illustre le schéma général de ce principe.

2 Utilisation des modèles MVL et MVG

2.1 Normalisation d'image par NSHP-HMM

Le NSHP-HMM possède une hauteur d'analyse (taille des colonnes observées) fixe. Par conséquent toutes les images doivent être normalisées à cette hauteur. Cette taille est identique pour tous les NSHP-HMM correspondant aux différentes classes de mots. Les échantillons sont des images binaires, 0 correspondant à un pixel blanc et 1 à un pixel noir. La normalisation en longueur de l'image s'appuie sur l'algorithme de Viterbi : étant donné un NSHP-HMM de mot et un échantillon, il fournit la meilleure répartition des états du NSHP-HMM sur les colonnes de l'image.
Soit $ObsSt[l]$ le numéro de l'état observant la colonne l. Soit N le nombre d'états du NSHP-HMM, et H sa hauteur d'analyse. H et L correspondent à la largeur et à la hauteur de l'image InI en entrée. L'image normalisée $NormI$ sera de dimension $N * H$, la hauteur étant fixée par celle du NSHP-HMM et la longueur par le nombre d'états du modèle. L'algorithme 1 décrit la méthode de normalisation. Intuitivement, elle correspond à calculer la densité de pixels observés par chaque état à chaque position verticale dans l'image; lorsqu'un état n'observe aucune colonne, la colonne correspondante de l'image normalisée est considérée comme blanche (pixels à 0). La figure 2 illustre le principe de normalisation et son résultat. Le niveau de gris de l'image destination représente la densité de pixels noirs observés par l'état correspondant à la colonne destination, à la position verticale considérée.
Deux informations supplémentaires sont ajoutées à chaque colonne de l'image normalisée :

1. la densité (proportion dans l'image) de colonnes observées par l'état correspondant,
2. la probabilité moyenne de ces colonnes pour cet état.

Ces paramètres ajoutent deux lignes à l'image normalisée.

2.2 Apprentissage des NSHP-HMM et des SVM

Suivant le principe exposé en §.1, le NSHP-HMM est utilisé pour la normalisation d'images et le SVM pour leur classification. Dans ce travail, ces deux modèles sont utilisés selon une approche "modèle discriminant". Par conséquent, un NSHP-HMM et un SVM sont associés à chaque classe de mot, le NSHP-HMM étant utilisé pour normaliser les images avant présentation au SVM. Ces deux modèles nécessitent une phase d'apprentissage.
L'entrainement des NSHP-HMM se fait par apprentissage croisé, comme décrit en [2]. Une fois entrainés, les NSHP-HMM sont utilisés pour normaliser les bases d'échantillons. Le SVM correspondant à chaque NSHP-HMM est alors en-

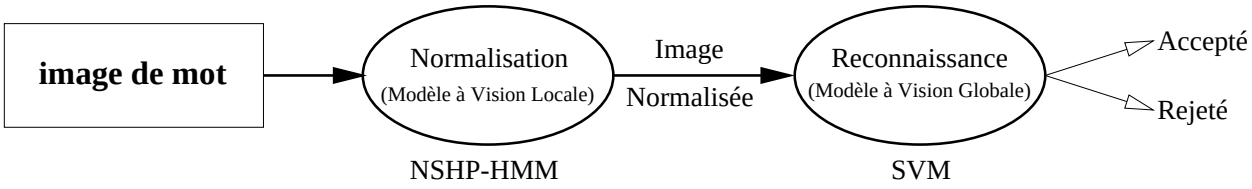

FIG. 1 – *Schéma général du système*

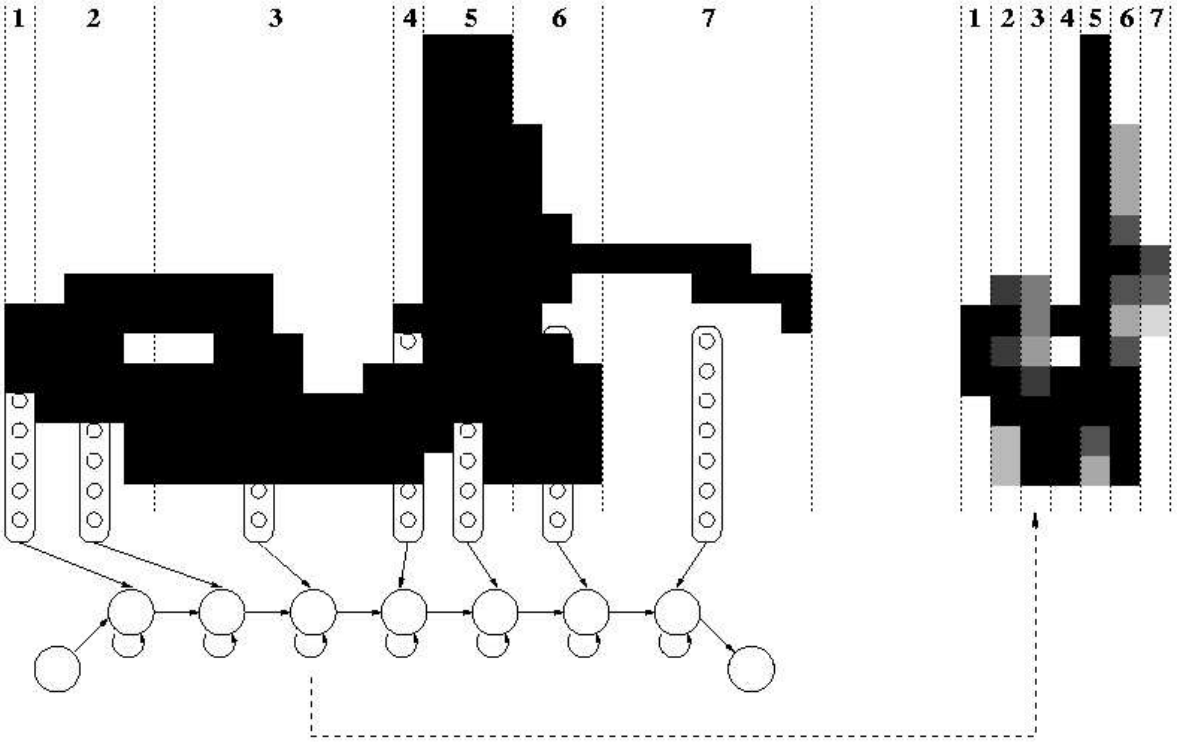

FIG. 2 – *Normalisation du mot "et" par le* NSHP-HMM *correspondant à cette classe. Le niveau de gris correspond à la densité de points noirs observés à cette position par l'état correspondant à la colonne de l'image normalisée.*

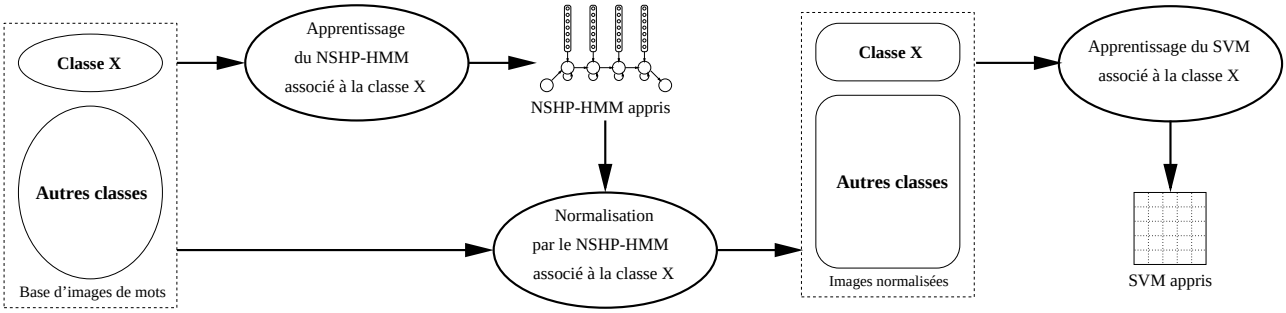

FIG. 3 – *Phases d'apprentissages pour une classe*

Algorithm 1 Normalisation d'une image binaire par un modèle de mot

Require: une image binaire $InI[][]$ et une séquence d'états $ObsSt[]$.

Ensure: une image normalisée $NormI[][]$

1: {Initialisation : $NbSt[]$ et $NormI[][]$ sont fixés à 0}
2: {Normalisation}
3: {Comptage du nombre de colonnes observées par chaque état}
4: **for** $l = 1$ to L **do**
5: $NbSt[ObsSt[l]] = NbSt[ObsSt[l]] + 1$
6: **end for**
7: {Normalise l'image en entrée}
8: **for** $l = 1$ to L **do**
9: **for** $h = 1$ to H **do**
10: {On ajoute le pixel courant à l'image normalisée et on divise par le nombre total de pixels à la hauteur h. Rappel : un pixel noir vaut 1, un pixel blanc vaut 0}
11: $NormI[h][ObsSt[l]] = NormI[h][ObsSt[l]] + InI[h][l] \ / \ NbSt[ObsSt[l]]$
12: **end for**
13: **end for**

trainé à séparer les images normalisées par ce dernier selon qu'elles appartiennent à sa classe ou non. La figure 3 illustre les différentes étapes de l'apprentissage pour une classe de mot.

2.3 Reconnaissance

Le processus global de reconnaissance est illustré par la figure 1. Un SVM renvoit en sortie la distance entre l'hyperplan séparateur et l'échantillon analysé. Une distance positive indique que l'image appartient à la classe, tandis qu'une distance négative indique un rejet. Cette distance n'est pas normalisée, donc ne peut être utilisée comme probabilité. Pour chaque image à reconnaitre, les SVM associés aux différentes classe de mot renvoient une distance. Les sorties des SVM peuvent être utilisées de trois manières simples :

- La première méthode consiste à classer les images pour lesquelles un seul SVM accepte l'appartenance, et à rejeter toutes les autres. Cette méthode est la plus juste mathématiquement parlant, puisque les distances renvoyées par les différents SVM ne sont pas des probabilités et ne sont pas normalisées. Leur comparaison n'a donc pas de sens mathématique.
- La seconde méthode consiste à prendre comme classe d'appartenance celle pour laquelle le SVM correspondant renvoit la distance la plus élevée. L'avantage de cette méthode est de n'avoir pas de rejet, ce qui permet une comparaison plus simple avec d'autres approches. Cependant la comparaison des distances entre différentes classes n'a pas beaucoup de signification.
- La troisième méthode consiste à combiner les deux précédentes. Les échantillons n'ayant que des distances négatives ou nulles sont rejetés, et la discrimination se fait sur les distances dans le cas de plusieurs réponses positives.

Les prochains travaux pourront utiliser une normalisation des distances comme proposé en [4]. L'utilisation de SVM multiclasses [6] permettra également de simplifier la classification.

3 Résultats et discussion

3.1 Résultats

Nous avons appliqué notre concept à la reconnaissance de mots de montants de chèques francais. La base de données contient 10321 images de qualité industrielle, réparties dans 28 classes selon la table 1. Les NSHP-HMM associés aux classes prennent en compte les principales erreurs d'orthographe. $2/3$ des échantillons (6798 images) servent à l'apprentissage des NSHP-HMM et des SVM; le $1/3$ restant (3523 images) est utilisé pour les tests.
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 estimation

TAB. 1 – *Distribution des échantillons dans les différentes classes de mots*

Label	un	deux	trois	quatre	cinq	six	sept	huit	neuf
Samples	48	580	334	612	381	245	214	209	175
dix	onze	douze	treize	quatorze	quinze	seize	vingt	trente	quarante
376	37	67	35	37	60	45	558	154	170
cinquante	soixante	cent	mille	francs	et	centimes	frcs	cts	
224	271	1693	950	2096	302	198	97	153	

étant faite à partir de la largeur moyenne des lettres). La dimension d'entrée du SVM correspondant à la classe est déterminée d'après ces information, conformément à la section 2.1.

La table 2 montre les résultats de la classification par les NSHP-HMM en utilisant l'algorithme de Viterbi, ainsi que les résultats donnés par les trois méthodes proposées pour la reconnaissance par les SVM après normalisation par les NSHP-HMM. À titre de comparaison, nous avons également testé la reconnaissance par les SVM avec une normalisation linéaire des images. La normalisation linéaire donne aux images les mêmes dimensions que la normalisation markovienne. Les ensembles d'apprentissage et de test sont les mêmes.

Les résultats sont similaires pour les deux méthodes de normalisation, lorsqu'on considère la classification avec rejet uniquement. On constate un taux de rejet plus important pour la normalisation par NSHP-HMM. L'observation des distributions d'états sur les images a montré que le principal problème provient d'une mauvaise répartition des états du NSHP-HMM sur les colonnes des images. Pour confirmer ce point, ainsi que pour obtenir une meilleure évaluation de notre hypothèse de travail, un nouveau test a été effectué en considérant uniquement les échantillons bien classés par les NSHP-HMM lors de la reconnaissance. A priori les répartitions d'états devraient être de meilleure qualité pour ces échantillons. La table 3 détaille les résultats de ce test.

Les nouveaux résultats restent similaires pour les deux méthodes de normalisation, mais les taux de rejets sont plus proches. Cela tend à confirmer l'importance des mauvaises répartitions d'états dans les résultats des premiers tests. On observe également une nette amélioration des résultats pour la classification basée sur les distances, ce qui confirme qu'une bonne étape de normalisation améliore notablement les résultats. Bien que les résultats obtenus ne tranchent pas clairement en faveur de la normalisation par les NSHP-HMM, cette voie semble prometteuse. Différents points vont être discutés sur ces résultats.

3.2 Discussion

L'un des principaux problèmes rencontrés dans ce travail est le manque d'échantillons dans plusieurs classes de mots. Cela conduit à un mauvais apprentissage des NSHP-HMM, qui ont alors des difficultés à répartir correctement leurs états sur les images. L'étape de normalisation dépendant de cette répartition, la qualité des images normalisées sera plus faible, ce qui conduit à une dégradation des performances des SVM tant en apprentissage qu'en reconnaissance. Une validation correcte de l'approche proposée nécessitera une base de mots plus complète.

Pour des questions de temps et de place, la puissance des NSHP-HMM n'a pas été entièrement utilisée. En effet, il est possible d'utiliser différentes symétries des images pour améliorer considérablement les résultats de ces modèles [10]. Le principe des symétries est de considérer les quatre images obtenues de l'original par combinaisons des flips vertical et horizontal de l'image. Un NSHP-HMM est créé pour chacune de ces images. Le score total de l'image analysée est le produit des scores obtenus par chaque modèle pour chaque symétrie. La table 4 montre l'impact des symétries sur les performances des NSHP-HMM. L'amélioration des résultats provient intuitivement du fait que la prise en compte des répartitions d'états sur les différentes symétries permet de "rattrapper" les erreurs de répartition d'états pouvant survenir pour une symétrie donnée.

Ces symétries sont utilisables avec les SVM de deux manières différentes :

- une première solution consiste à créer un SVM pour chaque symétrie, et combiner les résultats des quatres SVM pour l'acceptation ou le rejet de l'image.
- une seconde solution consiste à utiliser les répartition d'états obtenues pour les quatres NSHP-HMM et créer l'image normalisée à partir de ces quatre répartitions. Les résultats de la combinaison des quatre NSHP-HMM étant nettement meilleurs que ceux pour un seul NSHP-HMM, on peut s'attendre à ce que le mixage des quatre images normalisées soit meilleur qu'une image normalisée par un seul NSHP-HMM. Le SVM sera alors plus apte à classifier correctement l'image normalisée.

En ce qui concerne la normalisation linéaire, considérer les quatre symétries n'aura aucun effet. En effet, cette normalisation étant linéaire, cela revient à considérer les différentes symétries de l'image normalisée, ce qui n'apporte

TAB. 2 – *Scores de reconnaissance pour différentes approches*

Normalisation	Reconnaissance	Top 1	Top 2	Top 3	Top 5	Rejet
	NSHP-HMM	67.41%	79.05%	84.19%	89.81%	0.00%
NSHP-HMM	SVM+distance	72.64%	82.26%	86.57%	90.60%	0.00%
	SVM+rejet	92.64%	95.63%	97.20%	98.15%	40.25%
	SVM+dist+rejet	91.02%	93.22%	94.08%	94.57%	36.76%
Lineaire	SVM+distance	76.30%	85.35%	89.50%	93.50%	0.00%
	SVM+rejet	91.61%	95.20%	96.54%	97.79%	31.96%
	SVM+dist+rejet	90.35%	92.23%	93.19%	93.92%	29.09%

TAB. 3 – *Scores de reconnaissance sur les "bons" échantillons*

Normalisation	Reconnaissance	Top 1	Top 2	Top 3	Top 5	Rejet
NSHP-HMM	SVM+distance	85.14%	92.63%	95.12%	97.09%	0.00%
	SVM+rejet	96.59%	98.35%	98.98%	99.43%	25.94%
	SVM+dist+rejet	95.08%	98.22%	98.97%	99.41%	22.11%
Lineaire	SVM+distance	85.01%	92.46%	95.33%	97.35%	0.00%
	SVM+rejet	95.87%	98.46%	99.01%	99.45%	23.54%
	SVM+dist+rejet	94.73%	98.35%	98.94%	99.41%	20.84%

TAB. 4 – *Effet des symétries et des algorithmes sur les résultats des* NSHP-HMM

Recognition	Top 1	Top 2	Top 3	Top 5
Viterbi 1 symétrie	67.41%	79.05%	84.19%	89.81%
Viterbi 4 symétries	75.12%	85.58%	89.29%	92.83%
Baum-Welch 1 symétrie	70.93%	81.44%	85.97%	90.76%
Baum-Welch 4 symétries	76.40%	86.03%	89.20%	92.97%

aucune nouvelle information. Les résultats seront donc strictement identiques pour le SVM. En effet, les SVM étant basés sur un produit scalaire, ils sont insensibles aux symétries de l'espace d'entrée.

Si l'on prend en compte ces points, les résultats sont encourageants. Les résultats actuels sont similaires entre la normalisation par NSHP-HMM et la normalisation linéaire, avec un peu plus de rejet pour la normalisation markovienne pour une performance légèrement supérieure. Mais la normalisation markovienne est pénalisée par un manque d'échantillons, et sa puissance n'a pas été entièrement exploitée. L'amélioration de l'apprentissage est déterminante pour la qualité de la normalisation, et donc pour les résultats globaux. La comparaison des scores obtenus entre les deux séries de tests, principalement pour la classification sur les distances seulement, donne une idée des gains potentiels avec une meilleure qualité de normalisation.

3.3 Perspectives

La comparaison des résultats entre les approches linéaire et markovienne est relativement complexe à effectuer. En effet, les taux de rejet sont important et différents pour les deux approches, et les scores de reconnaissance interviennent sur le reste des échantillons. Notre analyse des résultats se base non pas sur la comparaison des taux de rejet et de reconnaissance pour les deux méthodes, mais sur l'amélioration des résultats (baisse du rejet et augmentation des scores) lorsque l'on "améliore" la normalisation (ici en considérant les images correctement classifiées par les NSHP-HMM, que l'on espère alors correctement normalisées). Si l'amélioration des scores pour les deux normalisation est relativement similaire, il n'en va pas de même des taux de rejet qui baissent beaucoup plus fortement pour la normalisation markovienne que pour la normalisation linéaire. Nous pensons que cette variation plus importante reflète l'intérêt d'une normalisation par MVL lorsqu'elle est correctement effectuée. Pour diverses raisons exposées cidessous, il est possible d'améliorer la qualité de la normalisation par les NSHP-HMM, et donc de valider notre idée.

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 [2]. En utilisant ces informations, il est possible d'utiliser les SVM 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. Il est intéressant de noter que les résultats présentés pour les NSHP-HMM sont déjà basés sur une approche analytique, et seraient améliorés en utilisant une approche globale à la place.

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* plutot que *Viterbi* pourrait améliorer cette étape (table 4). 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. Il serait intéressant d'affiner encore cette normalisation en descendant au niveau du pixel lui-même, par exemple en utilisant un PHMM. 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.

Il est à noter que la dépendance forte créée par le champ de Markov entre les colonnes observées peut jouer en défaveur du NSHP-HMM lors de la normalisation. Les travaux futurs devront comparer l'intérêt de la rigidité des MVL pour la normalisation.

Un défaut important des SVM est leur sortie qui ne peut pas être utilisée comme probabilité. L'utilisation de RN pour le MVG permettrait de dépasser ce problème. Le second défaut des SVM est lié à leur nature. La recherche de l'hyperplan séparateur s'appuie sur un produit scalaire entre vecteurs de données sans prendre en compte réellement la structure de ces données. Cela ne permet pas de faire de véritables corrélations entre parties d'une forme. Un troisième défaut provenant du produit scalaire est le poids accordé à chaque coordonnée du vecteur d'entrée : si les valeurs ne sont pas d'ordre similaire, leur influence sera variable, ce qui nécessite d'ajouter une normalisation des coordonnées. Cela a une influence particulièrement sur les informations de probabilité moyenne des colonnes, qui sont très petites et ont donc moins d'impact que les niveaux de gris. Ce problème n'apparait pas dans les RN car chaque entrée est pondérée, donc ces valeurs peuvent être automatiquement normalisées. Ces inconvénients renforcent l'intérêt des RN face aux SVM comme MVG.

Notre approche se décompose en deux parties. La première peut être utilisée comme préclassification 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 SVM pour déterminer la meilleure répartition d'états. De manière plus générale, les informations fournies par le MVG pourraient être utilisées pour améliorer la qualité de la normalisation par les MVL.

4 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 le SVM. La normalisation proposée s'appuie sur la répartition fournie par l'algorithme de *Viterbi* des états du NSHP-HMM sur l'image analysée. L'amélioration nette des résultats lorsqu'on considère les échantillons les plus correctement normalisés appuie l'intérêt de cette approche.
Le principal problème rencontré est le manque d'échantillons qui diminue les performances du NSHP-HMM et donc du système dans l'ensemble. D'autres tests sont nécessaires dans de meilleures conditions pour valider correctement cette approche. Les défauts des SVM rendraient également utiles leur remplacement par des RN.
Le schéma de travail proposé est très général, et de nombreux travaux sont à effectuer pour explorer toutes les possibilités d'une telle approche.

Références

[1] Rolf Bippus, *1-dimensonal and pseudo2-dimensional hmms for the recognition of german literal amounts*, Fourth International Conference on Document Analysis and Recognition (ICDAR'97) (Ulm, Germany), Aug. 1997.

[2] C. Choisy and A. Belaid, *Analytic Word Recognition Without Segmentation Based on Markov Random Fields*, Seventh International Workshop on Frontiers in Handwriting Recognition (IWFHR7) (Amsterdam), Sept. 2000.

[3] 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 (Rouen), juillet 1994, pp. 99–108.

[4] L. Vuurpijl F. Wang and L. Schomaker, *Support Vector Machines for the Classification of Western Handwritten Capitals*, Proc. IWFHR-7 (Amsterdam), 2000, pp. 167–176.

[5] M. Gilloux, B. Lemarié, and M. Leroux, *A Hybrid Radial Basis Function Network/Hidden Markov Model Handwritten Word Recognition System*, Third International Conference on Document Analysis and Recognition (ICDAR'95) (Montréal), 1995, pp. 394–397.

[6] Yann Guermeur, Andre Elisseeff, and Helene Paugam-Moisy, *A new multi-class SVM based on a uniform convergence result*, International Joint Conference on Neural Networks (IJCNN'2000), vol. IV, juillet 2000, pp. 183–188.

[7] T. Joachims, *Making large-scale svm learning practical. advances in kernel methods - support vector learning, b. schölkopf and c. burges and a. smola (ed.)*, MIT-Press, 1999.

[8] A. Kundu, *Handwritten Word Recognition using Hidden Markov Model*, Handbook of Character Recognition and Document Image Analysis (Nancy) (H. Bunke and P.S.P. Wang, eds.), 1997, pp. 157–182.

[9] H. S. Park and S. W. Lee, *An HMMRF-Based Statistical Approach for Off-line Handwritten Character Recognition*, IEEE Proceedings of ICPR'96, vol. 2, 1996, pp. 320–324.

[10] G. Saon and A. Belaïd, *Off-line handwritten word recognition using a mixed hmm-mrf approach*, International Conference on Document Analysis and Recognition (ICDAR'97), vol. 1, pp. 118-122, Ulm, Germany, August, 1997.

[11] A. W. Senior and A. J. Robinson, *An Off-Line Cursive Handwriting Recognition System*, IEEE Transactions on Pattern Recognition and Machine Intelligence **20** (1998), no. 3, 308–321.

[12] J. C. Simon, O. Baret, and N. Gorski, *A System for the Recognition of Handwritten Literal amounts of checks*, Internal Association for Pattern Recognition Workshop on Document Analysis System (DAS'94), Kaiserlautern, Germany, September 1994, pp. 135–155.

[13] Vladimir N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag New York, Inc., 1995.