



HAL
open science

Localisation dynamique de mots-clés par NSHP-HMM : application et optimisations

Christophe Choisy

► **To cite this version:**

Christophe Choisy. Localisation dynamique de mots-clés par NSHP-HMM : application et optimisations. Sep 2006, pp.115-120. hal-00113520

HAL Id: hal-00113520

<https://hal.science/hal-00113520>

Submitted on 13 Nov 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.

Localisation dynamique de mots-clés par NSHP-HMM : application et optimisations

C. Choisy

ITESOFT

Parc d'Andron - Le Sequoia 30470 Aimargues France

Résumé : *Cet article présente un système de recherche de mots-clé basé sur le NSHP-HMM. Ce modèle permet de générer dynamiquement des modèles globaux de mots à partir de modèles de lettres, et travaille sans segmentation de l'écriture. Nous décrivons tout d'abord notre système, qui s'appuie sur une concurrence entre des modèles de mots et un modèle générique de l'écriture : cette approche est appliquée à la recherche de mots-clé dans du courrier manuscrit libre. Nous montrons ensuite comment diviser les temps d'exécutions par 4, d'une part grâce à un codage en virgule fixe particulièrement adapté au NSHP-HMM, d'autre part en désactivant dynamiquement les modèles de mots dont la complexité théorique diffère trop de la complexité réelle de l'image analysée. Les premiers résultats obtenus sont encourageants, en particulier pour une application document par document.*

Mots-clés : NSHP-HMM, modèle générique, VITERBI, virgule fixe, longueur naturelle.

1 Introduction

Ces dernières années, les travaux en reconnaissance de l'écriture manuscrite hors-ligne ont permis des avancées notables pour des applications à vocabulaire ciblé de petite taille [MOR 01, LEE 05], ou de grande taille [VIN 03a, SCH 05]. Les vocabulaires réduits peuvent être efficacement traités par des modèles de mots en concurrence : leur faible taille permet d'utiliser cette approche en un temps très correct, et limite les confusions possibles. Les grands vocabulaires nécessitent quand à eux une mise en oeuvre plus complexe, réduisant le vocabulaire en cours de traitement, afin de conserver des temps d'exécution raisonnables.

L'avantage d'un vocabulaire réduit réside donc dans sa simplicité et sa rapidité de traitement, mais son inconvénient réside dans sa taille et sa spécialisation qui limitent fortement le champ d'application. A contrario, un vocabulaire large permet de s'adapter à un grand nombre d'applications, mais nécessite une mise en oeuvre complexe, comparative-ment coûteuse en temps, et présente beaucoup plus de risques de confusion ; de plus, les mots recherchés doivent nécessairement faire partie du vocabulaire traité [KOE 03a].

Le travail présenté répond à un besoin actuel de tri automatique de courriers manuscrits. Ce tri est basé sur un ensemble de mots-clés spécifiques aux services concernés. Dans ces conditions, considérer tout le vocabulaire du courrier est sans intérêt ; de plus ceci est impossible dans des courriers libres du fait de la présence de noms propres, termes techniques, etc. [VIN 03a]. Cependant le vocabulaire

des mots-clés ne doit pas être figé, sous peine de nécessiter un apprentissage spécifique pour chaque application. Les solutions proposées jusqu'ici pour le manuscrit se basent sur des informations globales des mots recherchés, les rendant spécifiques à l'application [KOL 97, BUR 98] ; de plus, ces travaux se focalisent sur la localisation du mot-clé, sans notion de rejet en cas d'absence.

Nous proposons donc une approche basée sur un vocabulaire réduit mais dynamique. De ce fait, nous conservons les avantages de rapidité et de simplicité des petits vocabulaires, tout en assurant la flexibilité des grands vocabulaires grâce à une création dynamique des mots à reconnaître.

Ce travail est basé sur le NSHP-HMM, modèle qui a fait ses preuves dans le cadre du vocabulaire réduit des mots de montants de chèques [SAO 97]. Ses évolutions ont permis un passage à l'analytique sans segmentation [CHO 02], autorisant la génération à l'envie de modèles globaux de mots : cette dynamique convient tout à fait au problème traité. Dans la première section, nous décrivons donc la mise en oeuvre du système de recherche de mots-clés, les adaptations nécessaires pour pallier les limites d'un vocabulaire figé, et les premiers résultats obtenus en recherche de mots-clé dans des courriers manuscrits.

La performance d'un tel système s'évalue également en fonction des temps de traitement. Dans la deuxième section, nous proposons deux méthodes permettant de réduire considérablement le coût de traitement. La première est basée sur une optimisation du codage des paramètres du modèle, qui prend tout son intérêt face à la combinatoire du NSHP-HMM. La seconde consiste à désactiver dynamiquement les modèles des mots en fonction de critères externes, en l'occurrence l'adéquation de la longueur naturelle de l'image avec celle théorique du mot recherché : cette approche a un double impact, sur les temps et sur les résultats.

Enfin, la dernière section présente nos conclusions et nos perspectives quand à ce système.

2 Utilisation du NSHP-HMM pour la recherche de mots-clé dans du courrier manuscrit libre

2.1 Rappels sur le NSHP-HMM

Le NSHP-HMM a déjà fait l'objet de plusieurs publications [SAO 97, CHO 02], aussi nous ne rappelons l'essentiel que brièvement. Ce modèle hybride un HMM et un champ

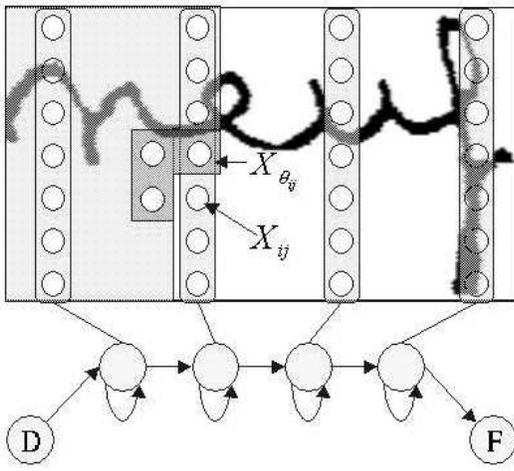


FIG. 1. NSHP-HMM : base HMM pour les colonnes et MRF considérant un voisinage pour les pixel

de Markov (MRF) ; il travaille directement sur des images binaires, qu'il analyse colonne par colonne pour la partie HMM, et pixel par pixel en considérant un voisinage pour la partie MRF.

La figure 1 décrit l'analyse d'une image par un NSHP-HMM, et l'équation 1 rappelle la formule d'estimation de la probabilité BAUM-WELCH d'une image de hauteur H et de longueur T [SAO 97] : le MRF prend en compte pour chaque pixel X_{ij} un voisinage de points $X_{\theta_{ij}}$, qui détermine sa probabilité conditionnelle de couleur.

$$P(X) = \sum_Q \prod_{j=1}^T \left[P(q_j | q_{j-1}) \prod_{i=1}^H P(X_{ij} | X_{\theta_{ij}}, q_j) \right] \quad (1)$$

Sous cette forme non factorisée, l'équation 1 montre la différence de complexité entre le NSHP-HMM et un HMM classique : la complexité calculatoire en terme de multiplications est en $O(HT)$, contre $O(T)$ pour le HMM classique ; Le facteur d'augmentation correspond à la partie MRF, qui estime la vraisemblance d'une colonne par le produit des H probabilités contextuelles de ses pixels.

Nous rappelons que la taille du MRF définit la hauteur H des colonnes analysées, imposant de fait une normalisation des images analysées : celles-ci subissent donc une normalisation différentielle, destinée à recentrer la zone des minuscules. L'estimation de cette bande centrale de l'écriture se fait par seuillage d'une combinaison des histogrammes de projection et de longueur naturelle.

2.2 Approche proposée

Dans [CHO 02] il est décrit comment générer, à partir de NSHP-HMM de lettres et d'un modèle d'enchaînement, des NSHP-HMM globaux de mots capables d'analyser une image sans segmentation. Ce principe a été appliqué à la reconnaissance de mots de montants de chèques : dans ce cas, le vocabulaire est connu et limité, et la reconnaissance se base sur la compétition entre modèles, la plus forte probabilité désignant la classe de l'image.

Dans le cas d'un texte libre, le vocabulaire analysé est sinon illimité, du moins potentiellement très grand. Il

est donc clair qu'une approche compétitive ne peut être utilisée pour la recherche de mots-clé : pour une image ne ressemblant à aucun mot cherché, la comparaison des réponses des modèles n'a pas de sens.

Pour résoudre ce problème, nous proposons la création d'un modèle générique représentant l'ensemble de l'univers écrit : son rôle est d'étalonner les scores de chaque modèle de mot. La compétition ne se fait donc plus entre les modèles de mots, mais entre chaque mot et le modèle générique : plus la différence entre le score d'un mot et le score du modèle générique est élevée, plus la réponse est pertinente et fiable.

L'idée intuitive sous-jacente est de simuler une compétition entre le modèle d'un mot recherché et le modèle réel de l'image analysée. Si l'on a un modèle générique de l'ensemble de l'écrit nous pouvons faire l'hypothèse classique que celui-ci s'aligne au mieux sur l'image, autrement dit que son comportement simule celui du modèle réel. La compétition entre le modèle générique et un modèle de mot prend alors tout son sens, le modèle générique servant de mesure-étalon pour le score de chaque modèle de mot recherché.

Diverses voies ont été explorées pour générer le modèle générique. L'idée intuitive de mettre en boucle tous les modèles de lettres, ou des sous-ensembles selon les hampes et jambages, a été abandonnée à cause de son coût de traitement important. Des résultats sensiblement équivalents ont en effet été obtenus avec un unique caractère générique, appris sur toutes les images de mots.

Cet apprentissage peut être effectué en faisant boucler le caractère générique sur les images : cela lui permet de se focaliser sur les éléments structurants de l'écriture ; il peut également se faire en remplaçant les lettres par le modèle générique : ceci s'apparente à une interpolation des modèles de lettres. Les résultats sont similaires dans les deux cas, mais la seconde approche est plus rapide en apprentissage, tandis que la première ouvre certaines perspectives discutées en fin d'article.

En reconnaissance, le modèle générique est appliqué en boucle sur les images. Après normalisation, le score de chaque mot du vocabulaire est comparé au score du modèle générique, et un seuillage de la différence des probabilités permet de sélectionner le mot gagnant, ou de rejeter l'image comme n'étant pas un mot-clé.

2.3 Application

Nous avons appliqué ce système à la recherche de mots dans du courrier manuscrit français. 1522 documents ont été étiquetés manuellement ; tous sont d'auteurs différents. La base de test est formée de 508 courriers contenant 22738 "mots". L'apprentissage se fait sur les courriers restants ; cependant, afin de favoriser la détection de la bande centrale, seuls les 26350 "mots" de plus de 4 caractères ont été retenus. Les lettres accentuées sont regroupées avec leur homologue sans accent, pour éviter la gestion des différentes variantes orthographiques consécutives et la raréfaction des échantillons. Les tests ont été réalisés sur un Pentium®4 ca-

dencé à 1.4Ghz.

L'apprentissage consiste en 160 passes de réestimation des lettres et du caractère générique. Ce choix est basé sur de précédentes expérimentations, où nous avons observé que l'apprentissage converge en permanence lorsque les modèles reproduisent correctement la syntaxe des images de mots. À 80 passes environ, les résultats commencent à être de bonne qualité; 160 passes nous assurent donc une modélisation correcte des lettres. Lorsque les documents étiquetés seront en nombre suffisant pour créer une base de validation, il serait intéressant de vérifier que tous les caractères sont correctement représentés et qu'il n'y a pas de surapprentissage : nos précédentes expériences, ainsi que les 26350 mots de la base d'apprentissage, nous ont amenées à faire l'hypothèse que non pour ce travail.

Nous avons choisi un vocabulaire de 15 mots représentatifs des différences de taille et de présence. Les confusions dues à des variantes orthographiques proches gardant le sens du mot cherché ont été tolérées : pour un tri automatique, le sens est plus prégnant que la forme même. Chaque mot possède un seuil de rejet propre ; ce seuil a été relevé sur la base de test, la quantité de documents étiquetés étant trop faible pour créer une base de validation.

La table 1 résume les taux de détection de chaque mot pour un seuillage idéal rejetant toutes les erreurs. Notons que les résultats portent uniquement sur la bonne détection des mots-clé, et n'intègrent pas la notion de rejet correct des mots : cet aspect reste à étudier, et nécessite un second seuil pour chaque mot.

exoneration	changement	adresse	domicile
29.79%	28.57%	37.25%	55.56%
prelevement	reclamation	telephone	televiseur
13.33%	25%	50%	36.90%
redevance	compte	cheque	vous
41.52%	48.89%	58.82%	4.79%
service	Monsieur	Madame	
18.67%	33.61%	57.21%	

TAB. 1. Taux de repérage des différents mots-clé

La moyenne des scores est de 35.994%, et la moyenne pondérée par le nombre de mots est de 28.41%. Si l'on retire le mot 'vous', volontairement gardé pour sa spécificité de taille et de fréquence pénalisantes, la moyenne est de 38.22%, et la moyenne pondérée est de 39.67%.

Ces résultats sont variables. Pour le mot 'vous', la principale cause d'échec est la taille réduite du mot qui conduit à une mauvaise détection de la bande centrale : les variations dans le corps des minuscules introduisent des hampes ou jambages artificiels qui faussent la reconnaissance. Le modèle de 'prelevement' se trompe majoritairement sur des images du mot 'paiement' ; cette erreur provient vraisemblablement de deux points : d'une part, la modélisation cumulée des 'e' avec ou sans accent introduit un biais dans les probabilités de chacun de ces caractères ; d'autre part, la détection de la bande centrale est parfois faussée par la présence des

accents, qui provoquent alors un écrasement de l'information dans les différentes zones de l'écriture lors de la normalisation différentielle de l'image.

Nous avons observé globalement que la variabilité des caractères dans la bande centrale ainsi que la bonne détection de celle-ci sont des critères importants pour l'efficacité du NSHP-HMM, et expliquent une grande part des erreurs lorsque les formes globales des mots sont similaires. Le système actuel est basé sur des prétraitements minimaux et légers : l'ajout de prétraitements plus complexes, comme le suivi du corps des minuscules, pourrait donc améliorer ces résultats.

Pour se faire un idée plus précise du potentiel de l'approche proposée, il est intéressant de regarder un autre résultat : celui de la présence du mot recherché en première(s) position(s) dans chaque document. Ce score est reporté dans la table 2.

exoneration	changement	adresse	domicile
72.34%	92.86%	79.41%	94.44%
prelevement	reclamation	telephone	televiseur
100%	75%	91.67%	82.14%
redevance	compte	cheque	vous
92.42%	97.78%	96.08%	77.07%
service	Monsieur	Madame	
77.33%	92.21%	92.09%	

TAB. 2. Taux de présence des mots en première position dans un document

Alors que la reconnaissance des mots donne dans l'absolu des résultats moyens, on peut constater que dans la majorité des cas, lorsque des images du mot sont présentes dans le document, leurs scores les placent en premières positions des réponses. Cela montre qu'avec une adaptation au document, que l'on pourrait assimiler à une adaptation au scripteur, il est possible d'améliorer nettement les résultats du système.

3 Accélération du système

La rapidité d'un système de reconnaissance de mots tient principalement à trois paramètres : la complexité des données traitées, la complexité du système en terme de calculs et de mise en oeuvre, et la quantité de données à traiter effectivement.

Des travaux précédents sur le NSHP-HMM ont montré qu'une réduction de la taille de l'image analysée pénalise fortement la qualité des résultats : cette voie n'a donc pas été retenue. Notre attention s'est portée sur les deux derniers paramètres, d'une part en réduisant la complexité calculatoire du système pour le processeur, et d'autre part en invalidant les modèles les moins pertinents selon les images traitées.

3.1 Réduction des temps de calcul

La réduction du coût des calculs s'est faite en plusieurs étapes. Les premières sont tout à fait classiques : passage de la reconnaissance BAUM-WELCH à l'analyse par VITERBI, puis passage aux log-probabilités permettant de passer de produits de réels à des sommes de réels. L'équation 2 montre

l'intérêt du log-VITERBI en comparaison avec l'équation 1, intérêt multiplié par la combinatoire propre au NSHP-HMM.

$$L_p(X) = \arg \max_Q \sum_{j=1}^T \left[L_p(q_j | q_{j-1}) + \sum_{i=1}^H L_p(X_{ij} | X_{\theta_{ij}}, q_j) \right] \quad (2)$$

Cependant, la nature du NSHP-HMM permet de contenir ses paramètres dans des plages de précision relativement faibles ; nous avons donc profité de cet état de fait pour appliquer un codage en virgule fixe de ces paramètres, permettant alors de passer de sommes de réels à des sommes d'entiers. La table 3 résume les durées moyennes de reconnaissance sur 10 exécutions en fonction des différentes approches, ainsi que les proportions de temps.

méthode	BW	V	LogV	ILogV
durée	1069.07s	951.72s	794.35s	468.56s
ms/mot	47.02ms	41.86ms	34.93ms	20.61ms
mots/s	21.27	23.89	28.62	48.53
% de BW	100%	89.02%	74.30%	43.83%
% de V	112.33%	100%	83.46%	49.23%
% de LogV	134.58%	119.81%	100%	58.99%
% d'ILogV	228.16%	203.12%	169.53%	100%

TAB. 3. Temps de calcul et proportion en fonction des méthodes d'analyse : BAUM-WELCH, VITERBI, log-VITERBI, int-log-VITERBI

En moyenne, le système permet de traiter 48.53 images de mots par seconde dans sa version la plus rapide, soit 20.61ms par image. En considérant le vocabulaire de 15 mots recherchés et l'analyse par le modèle générique, le coût moyen pour un modèle sur une image est de 1.288ms.

Les résultats montrent clairement l'intérêt du codage en virgule fixe pour le NSHP-HMM : si le passage aux logarithmes permet un gain proche de 17% de temps par rapport à VITERBI, le codage entier en économise plus de 50%. De plus, l'impact sur les résultats est négligeable : applicativement, l'ordre des réponses est le même pour toutes les approches ; numériquement, la différence de probabilités mesurée entre VITERBI et log-VITERBI est en moyenne de l'ordre de 10^{-7} , et la différence avec int-log-VITERBI est au maximum de l'ordre de 10^{-5} , et en moyenne de l'ordre de 10^{-6} .

Il est à noter que le gain du passage de BAUM-WELCH à VITERBI est relativement faible (environ 11%). Après plusieurs séries de tests, il semblerait que sur les processeurs utilisés, comparer deux réels pour éviter une addition soit plus coûteux que faire l'addition. Ce phénomène est vraisemblablement lié au fait que la génération 4 des processeurs d'Intel® a subi une diminution notable de performance en calculs réel par rapport à la génération 3, au profit des calculs entiers : cela accentue encore l'intérêt du passage en virgule fixe sur ces processeurs.

3.2 Désactivation dynamique des modèles non pertinents

L'utilisation de critères globaux pour filtrer un dictionnaire en fonction de l'image analysée est courante. Ces cri-

tères peuvent être basés sur des caractéristiques structurelles, morphologiques, etc. [KOE 03a, VIN 03b].

Nous avons choisi de reprendre la longueur naturelle, mesure classique de la complexité d'un mot, déjà extraite lors des prétraitements pour la localisation de la bande centrale (cf. § 2.1). Son avantage est de faire très peu d'hypothèses sur la forme du mot : seule la complexité graphique de la bande centrale est prise en compte, ce qui évite les interprétations locales de l'image telles que hampes, jambages, etc.

Lors de l'apprentissage, la contribution moyenne de chaque caractère à la longueur naturelle des mots est estimée approximativement en supposant une contribution égale des lettres au mot. Cette estimation est grossière, mais suffisante pour évaluer l'intérêt de cette mesure et du filtrage dans notre système.

En phase de reconnaissance, la longueur naturelle moyenne théorique de chaque mot du vocabulaire est reconstituée à partir de la contribution moyenne des lettres. Lors de l'analyse d'une image, sa longueur naturelle est comparée à celle théorique de chaque modèle : si la différence excède un certain seuil, supérieur ou inférieur, le modèle est désactivé et ne sera donc pas appliqué sur l'image.

Ce filtrage a été appliqué dans les conditions des précédents tests, en considérant les seuils optimaux pour chaque mot. Ces seuils englobent les variantes orthographiques tolérées précédemment car porteuses du sens du mot, limitant de fait la qualité du filtrage. Le temps de calcul pour la version en virgule fixe est de 218.78s, soit 46.69% du temps initial (468.56s) : le temps de traitement est donc divisé par plus de 2. Le traitement d'une image par les 15 modèles de mots et le modèle générique prend 9.622ms, et le coût moyen d'un modèle sur une image est de 0.6ms.

Les colonnes de pixels des images étant la base des calculs du NSHP-HMM, la table 4 indique pour chaque mot le taux de colonnes d'images analysées après filtrage.

exoneration	changement	adresse	domicile
23.27%	44.47%	62.23%	35.69%
prelevement	reclamation	telephone	televiseur
26.11%	21.66%	27.66%	48.23%
redevance	compte	cheque	vous
55.29%	47.44%	49.84%	55.67%
service	Monsieur	Madame	
57.10%	65.65%	51.13%	

TAB. 4. Taux de colonnes supprimées pour chaque mot

Au vu des résultats, il est clair que la longueur naturelle est efficace pour économiser de nombreux calculs. La longueur des labels des mots du vocabulaire a un impact fort sur le taux de filtrage : les mots les plus longs filtrent le plus d'images. Ce résultat est cohérent, car dans les courriers il existe peu de mots plus longs pouvant être englobés dans la limite supérieure de longueur naturelle, la limite inférieure filtrant quand à elle de nombreux bruits.

Les pourcentages élevés de 'Monsieur' et 'téléviseur' sont justifiés principalement par les variantes orthographiques tolérées, qui faussent nettement les limites retenues pour la longueur naturelle.

Il est intéressant de noter que ce filtrage a également un effet de bord sur les résultats. En effet, la longueur naturelle est une mesure de complexité indépendante de la forme générale du mot, et donc relativement orthogonale à la modélisation globale du NSHP-HMM. Par conséquent certaines confusions dues à des formes globales proches sont filtrées par la longueur naturelle, ce qui améliore les résultats du système. La table 5 reporte l'influence du filtrage sur le pourcentage du mot en première position dans les documents lorsqu'il est présent, similairement à la section précédente.

exoneration	changement	adresse	domicile
76.60%	92.86%	79.41%	100%
prelevement	reclamation	telephone	televiseur
100%	75%	100%	85.71%
redevance	compte	cheque	vous
92.78%	97.78%	98.04%	77.65%
service	Monsieur	Madame	
77.33%	92.42%	93.49%	

TAB. 5. Résultats de reconnaissance par document après filtrage

Pour 9 mots sur 15 le taux de présence en première position dans les documents a été amélioré. De manière similaire, nous avons constaté pour tous les mots que les scores de reconnaissance augmentent plus rapidement avec le rang de la réponse, tant par document que de manière absolue. Le filtrage permet donc d'éliminer des images dont la forme générale est suffisamment proche du mot pour créer une confusion, mais dont la complexité s'en éloigne notablement.

Notre approche de recherche de mots-clé permet donc de profiter facilement d'un élagage classique des travaux à large vocabulaire, sans en nécessiter la complexité de mise en oeuvre et de calculs. La rapidité de traitement et la souplesse de gestion nous conforte dans l'intérêt d'une approche à mots-clé pour la classification de documents manuscrits non contraints.

3.3 Comparaison avec les approches à large vocabulaire

La table 6 permet de comparer les résultats en termes de temps et de qualité par rapport à des approches à large vocabulaire.

Les temps de calcul des approches à large vocabulaire restent globalement trop élevés pour un traitement en temps réel. Seul [SCH 05] propose un temps raisonnable pour un dictionnaire de très grande taille ; cependant son champ d'application concerne les noms de rues allemands, le mot recherché doit appartenir au dictionnaire, sa machine est nettement plus rapide que les autres, et ses résultats sont nettement inférieurs aux autres approches à large vocabulaires : ils se rapprochent de ceux de notre système pour la reconnaissance de

référence	machine	taille dic.	ms/mot	score
[KOE 03a]	Ath. 1.1Ghz	85.1k	10700ms	71.03%
[KOE 04]	Ath. 1.1Ghz	10k	1992ms	81.06%
[KOE 04]	Ath. 1.1Ghz	40k	7516ms	73.23%
[KOE 04]	Ath. 1.1Ghz	80k	14460ms	68.65%
[SCH 05]	P4 2.8Ghz	478123	135.7ms	48.1%
[SCH 05]	P4 2.8Ghz	478123	145.2ms	51.4%
[§ 3.1]	P4 1.4Ghz	15 dyn.	20.61ms	table 1
[§ 3.2]	P4 1.4Ghz	15 dyn.	9.622ms	table 1

TAB. 6. Comparaison des temps de calcul par mot avec des approches à large vocabulaire

mots.

Malgré la différence de configuration machine, notre meilleure approche est plus de 14 fois plus rapide que le meilleur temps de [SCH 05], et nous profitons de la possibilité de rejeter les mots n'étant pas dans le dictionnaire. Notre système étant basé sur une représentation plate du dictionnaire, il est extrêmement simple à mettre en oeuvre, contrairement aux approches à large vocabulaire. De plus, notre approche permet d'utiliser avec efficacité des méthodes de filtrage du dictionnaire classiques des approches à vocabulaire large. En point final, [VIN 03a] rappelle que ces dernières ne peuvent résoudre le problème de lecture "pleine page" : or, c'est la condition sine-qua-non pour pouvoir localiser n'importe quel mot-clé.

Cette comparaison nous permet donc d'estimer nos résultats prometteurs, et conforte l'intérêt d'une approche à recherche de mots-clé par rapport à une lecture "pleine page" pour le tri automatique de documents.

4 Conclusion et perspectives

Cet article montre comment le NSHP-HMM peut être utilisé en recherche de mots-clé dans des documents manuscrits, ainsi que des méthodes permettant d'accroître notablement son efficacité en terme de temps de traitement.

Les premiers résultats obtenus sont encourageants, en particulier l'efficacité par document qui laisse augurer de nettes améliorations des taux de reconnaissance. Les techniques d'accélération du système permettent de diviser les temps de traitement par 4 par rapport à une approche basée sur un VITERBI classique, grâce à l'optimisation du codage du NSHP-HMM et à la désactivation dynamique des modèles de mots en fonction de leur inadéquation aux images analysées. La flexibilité et la rapidité de notre approche en font une alternative intéressante à une lecture "pleine page" à large vocabulaire, pour un tri automatique de courriers manuscrits libres.

Une perspective importante de ce travail réside dans l'adaptation au document. Le NSHP-HMM travaillant sans segmentation et directement sur les pixels, il est très difficile de catégoriser ses observations afin de spécialiser ses paramètres. Cependant, comme nous l'avons évoqué en première section, le modèle générique utilisé pour l'étalonnage des scores peut être entraîné de plusieurs manières, l'une d'entre elles étant en boucle sur les images. Cette méthode pourrait

être appliquée pour spécialiser ce modèle sur chaque document, en le réapprenant sur les images de complexité graphique suffisante. Cela permettrait alors d'adapter le score de référence au document, ou de mesurer la différence de qualité entre le modèle d'origine et le modèle spécialisé pour estimer le biais dû à la nature de l'écriture.

Une autre piste consiste à changer les critères d'apprentissage afin d'égaliser au maximum les probabilités normalisées des images. En lissant ces résultats, on peut espérer "égaliser" la qualité de reconnaissance entre tous les types d'écriture.

Concernant l'accélération du système, la démarche de filtrage des modèles est destinée à évoluer. En effet, la longueur naturelle est un argument uniquement filtrant, et ne contribue donc pas au score de reconnaissance des images. Nous nous orientons donc vers l'intégration de systèmes probabilistes plus légers que le NSHP-HMM, qui permettraient à la fois de désactiver les modèles non pertinents et d'influer sur les résultats en fonction de leur apport. Le support est actuellement en place, et l'effort se porte sur l'extraction d'informations pertinentes.

Parallèlement, nous avons évoqué l'estimation grossière de la contribution de chaque lettre à la longueur naturelle de l'écriture. Les gains apportés par son utilisation encouragent à affiner son estimation. Celle-ci pourrait se baser sur l'analyse de VITERBI des images apprises, avec la supposition classique que le modèle réel du mot va s'aligner correctement sur les informations, déterminant les limites "réelles" des lettres. D'autre part, l'estimation actuelle se fait sur la zone des minuscules, déterminée par une combinaison d'histogrammes. Cela introduit un biais sur l'estimation de la longueur naturelle seule : il semble donc nécessaire de réadapter la zone d'analyse afin de mieux cibler les éléments porteurs de complexité.

Enfin, certaines études complémentaires doivent être menées sur le système actuel. En particulier, l'intégration de la notion de rejet correct des images n'étant pas des mots-clé, permettrait d'évaluer le taux de classification correcte des images de mots et des documents. Ces deux points n'ont pas été abordés au cours du travail actuel : nous avons limité nos investigations à la qualité de la reconnaissance des mots ; mais dans un cadre applicatif la notion de bonne classification des documents présente un intérêt évident.

Cet aspect pourrait être complété par l'étude du rang des mots-clé dans les documents : le taux important de positionnement des mots-clé en premier rang des résultats par document permet d'envisager une application d'aide à la saisie, en supposant que les N premières réponses contiennent les mots-clé s'ils sont présents dans le document.

Références

- [BUR 98] BURL M., PERONA P., Using hierarchical shape models to spot keywords in cursive handwriting data, *Computer Vision and Pattern Recognition*, 1998.
- [CHO 02] CHOISY C., BELAID A., Cross-learning in Analytic Word Recognition Without Segmentation, *International Journal On Document Analysis and Recognition*, vol. 4, 2002, pp. 281-289.

- [KOE 03a] KOERICH A. L., SABOURIN R., SUEN C. Y., Large Vocabulary Off-Line Handwriting Recognition : A Survey, *Pattern Analysis and Applications*, vol. 6, n° 2, 2003, pp. 97-121.

- [KOE 04] KOERICH A. L., SABOURIN R., SUEN C. Y., Fast Two-Level HMM Decoding Algorithm for Large Vocabulary Handwriting Recognition, *9th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR)*, Tokyo, Japan, 2004.

- [KOL 97] KOLCZ A., ALSPECTOR J., AUGUSTEIJN M., CARLSON R., POPESCU G. V., Keyword Search In Handwritten Documents, *Proceedings of the Third International Workshop on Applications of Neural Networks to Telecommunications*, Melbourne, Australia, 1997, pp. 171-180.

- [LEE 05] LEE L. L., COELHO S. I., A Simple and Efficient Method for Global Handwritten Word Recognition Applied to Brazilian Bankchecks, *International Conference on Document Analysis and Recognition*, vol. 2, 2005, pp. 950-954.

- [MOR 01] MORITA M., EL-YACOUBI A., BORTOLOZZI F., SABOURIN R., SUEN C. Y., Handwritten Month Word Recognition on Brazilian Bank Cheques, *International Conference on Document Analysis and Recognition*, 2001.

- [SAO 97] SAON G., BELAÏD A., High Performance Unconstrained Word Recognition System Combining HMMs and Markov Random Field, *International Journal on Pattern Recognition and Artificial Intelligence, Special Issue on Automatic Bankcheck Processing*, vol. 11, 1997, pp. 771-788.

- [SCH 05] SCHAMBACH M.-P., Fast Script Word Recognition with Very Large Vocabulary, *International Conference on Document Analysis and Recognition*, vol. 1, 2005, pp. 9-13.

- [VIN 03a] VINCIARELLI, BENGIO, Offline Recognition of Large Vocabulary Cursive Handwritten Text, *International Conference on Document Analysis and Recognition*, 2003.

- [VIN 03b] VINCIARELLI A., Offline Cursive Handwriting : From Word To Text Recognition, IDIAP-RR n° 24, 2003, Dalle Molle Institute of Perceptual Artificial Intelligence.