



HAL
open science

Combinaison de classifieurs HMMs à fenêtres symétriques et asymétriques pour la reconnaissance de mots manuscrits arabes

R. El-Hajj M., C. Mokbel, Laurence Likforman-Sulem

► **To cite this version:**

R. El-Hajj M., C. Mokbel, Laurence Likforman-Sulem. Combinaison de classifieurs HMMs à fenêtres symétriques et asymétriques pour la reconnaissance de mots manuscrits arabes. Colloque International Francophone sur l'Écrit et le Document, Oct 2008, France. pp.115-120. hal-00334405

HAL Id: hal-00334405

<https://hal.science/hal-00334405>

Submitted on 26 Oct 2008

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.

Combinaison de classifieurs HMMs à fenêtres symétriques et asymétriques pour la reconnaissance de mots manuscrits arabes

Rami El-Hajj M.^{1,2}, Chafic Mokbel¹ et Laurence Likforman-Sulem²

¹Université de Balamand, Faculté de Génie
P.O.Box 100 Tripoli - LIBAN

GET-Ecole Nationale Supérieure des Télécommunications / TSI
46, rue Barrault, 75013 Paris - FRANCE

al-hajj@enst.fr Chafic.Mokbel@balamand.edu.lb likforman@telecom-paristech.fr

Résumé : Le problème adressé dans cet article est la reconnaissance automatique hors ligne des noms de villes arabes. Ces noms appartiennent à un dictionnaire d'environ de 1000 classes. Notre méthode prend en considération les inclinaisons ainsi que les positions erronées des points et des marques diacritiques possibles dans l'écriture qui sont des sources majeures d'erreurs pour la reconnaissance des mots. La méthode proposée est composée d'une étape de reconnaissance et d'une étape de combinaison en post-traitement. L'étape de reconnaissance est basée sur 3 classifieurs de type modèles de Markov cachés (HMMs) qui utilisent des fenêtres glissantes de différentes orientations. L'étape de combinaison fusionne les sorties produites par les HMMs pour choisir le mot candidat le plus approprié. Plusieurs stratégies de combinaisons ont été examinées avec différentes valeurs d'inclinaison des fenêtres, ainsi que des orientations symétriques et non symétriques. Les résultats obtenus sur la base de référence IFN/ENIT montrent que la combinaison des classifieurs est meilleure que le système simple incluant une étape de correction d'inclinaison d'écriture.

Mots-clés : Ecriture manuscrite arabe – Modèles de Markov cachés – Extraction de primitives – combinaison des classifieurs.

1 Introduction

La reconnaissance de l'écriture arabe a de nombreuses applications notamment en bureautique et tri automatique du courrier, et pour l'enregistrement et la vérification des chèques bancaires. L'écriture manuscrite arabe est naturellement cursive, difficile à segmenter, et présente une grande variabilité de forme et de style. Les caractères incluent des hampes et des jambages qui peuvent être plus ou moins inclinés. Les caractères d'un mot ou d'un pseudo mot (composante connectée d'un mot), sont ligaturés horizontalement sur la ligne d'écriture dite ligne de base. Des ligatures verticales de deux ou trois caractères sont aussi possibles. Les espaces

inter et intra-mots sont de longueurs variables. De plus, les mots contiennent des points et des marques diacritiques qui peuvent changer le sens d'un mot. La position de ces points et marques diacritiques doit être prise en considération par un système de reconnaissance automatique.

La capacité des approches HMMs à modéliser stochastiquement les séquences d'observations de longueurs variables et à traiter les distorsions non linéaires, ont conduit à plusieurs systèmes basés sur cette approche pour la reconnaissance de l'écriture arabe [MOH 96] [PEC 03] [LOR 05] [BEN 03] [BEN 00]. Une méthode de type HMM combine généralement les avantages des stratégies de reconnaissance holistiques (globales) et analytiques. En effet, l'approche analytique, basée sur la modélisation d'un mot comme concaténation des modèles des caractères, est adaptée pour des applications à grand vocabulaire ou à vocabulaire ouvert [KHO 03]. D'autre part, l'avantage d'une approche holistique est de pouvoir reconnaître les mots sans pré-segmentation des mots en ses caractères constituants. De fait, une segmentation implicite des mots en ses caractères est réalisée au cours du décodage lors de la reconnaissance. L'entraînement des modèles des lettres est accompli globalement au travers des images des mots et de leurs transcriptions, sans segmentation manuelle en caractères.

Dans cet article nous présentons nos résultats complets sur la combinaison des classifieurs à fenêtres orientées. Nos premiers résultats [ELH 07] ont montré une amélioration significative du taux de reconnaissance par l'utilisation de 3 classifieurs à fenêtres symétriques. Cette amélioration vient de la prise en compte par les fenêtres inclinées des chevauchements possibles des lettres adjacentes, et des positions erronées des points et des marques diacritiques possibles dans l'écriture. Ces problèmes ont été identifiés comme des sources majeures d'erreurs de reconnaissance au cours d'une étude préliminaire sur la performance de notre système de référence présenté dans [ELH 05]. Trois stratégies de

combinaison sont examinées et des valeurs variées d'inclinaisons des fenêtres glissantes sont comparées. Les inclinaisons des classifieurs orientées peuvent être soit symétriques, soit asymétriques. Les résultats d'expériences montrent aussi que la combinaison des classifieurs fournit des résultats meilleurs que ceux fournis par un seul système incluant une étape de correction d'inclinaison d'écriture en prétraitement.

L'article est organisé de la façon suivante: Dans la section 2, nous présentons le système de référence et l'architecture générale du système proposé. La troisième section décrit le système de combinaison intégrant l'information sur l'orientation dans l'écriture. La section 4 présente les expériences d'évaluation de la performance du système sur la base de données IFN/ENIT de noms de ville Tunisiens [PEC 02]. Finalement, des conclusions ainsi que des perspectives futures sont présentées dans la section 5.

2 Système de référence

Le système proposé (Fig. 1) consiste à combiner trois classifieurs HMMs au niveau de la décision. Ces classifieurs sont issus du système dit de référence [ELH 05] dont nous rappelons les caractéristiques.

Le classifieur de référence est basé sur des HMMs niveau caractères de type gauche-droite. Les vecteurs de caractéristiques sont extraits de droite à gauche sur les images binaires de mots par des fenêtres verticales glissantes et intercalées. La reconnaissance est basée sur un ensemble de caractéristiques reflétant la distribution de densité et les concavités locales des pixels d'écritures.

Une première version de ce système est décrite dans [ELH 05]. Ce système a obtenu de bons résultats à la

compétition organisée à ICDAR 05 [MAR 05]. Une première étape de prétraitement extrait les deux lignes de base sur chaque image de mot: ligne haute et ligne basse. L'image du mot est ensuite balayée de droite à gauche par une fenêtre glissante qui s'adapte en hauteur à celle du mot [SCH 96]. Dans chaque fenêtre est extrait un ensemble de 28 caractéristiques. Celles ci sont représentatives des densités et des configurations locales des pixels d'écriture. Un sous ensemble des caractéristiques sont liées à la position des lignes de base pour prendre en compte la présence des hampes, des jambages et des points diacritiques.

Le système de reconnaissance de référence modélise les mots et les caractères sous la forme de modèles de Markov cachés (HMMs). Le système est analytique : les modèles de mots sont construits par concaténation des modèles de caractères. Ceux ci sont de type droite-gauche à 4 états comme le montre la Figure 2. Les densités de probabilité des observations dans chaque état sont modélisées par un mélange de 3 gaussiennes. Au total, 167 modèles de caractères sont construits pour l'écriture arabe qui incluent les formes différentes de caractères suivant leur position dans le mot (au début, au milieu ou à la fin), les formes modifiées du fait de la présence de marques additionnelles telles que le "shadda" (ّ), des chiffres possibles dans les mots de la base IFN/ENIT, et des caractères ligaturés utilisés dans cette base. Un modèle espace est aussi inclus dans l'ensemble des modèles de caractères. Celui est toujours inséré entre deux mots, et il est inséré entre deux pseudo-mots quand il n'y a pas de risque de chevauchement entre le caractère en fin de sous mot et le caractère en début du sous mot suivant.

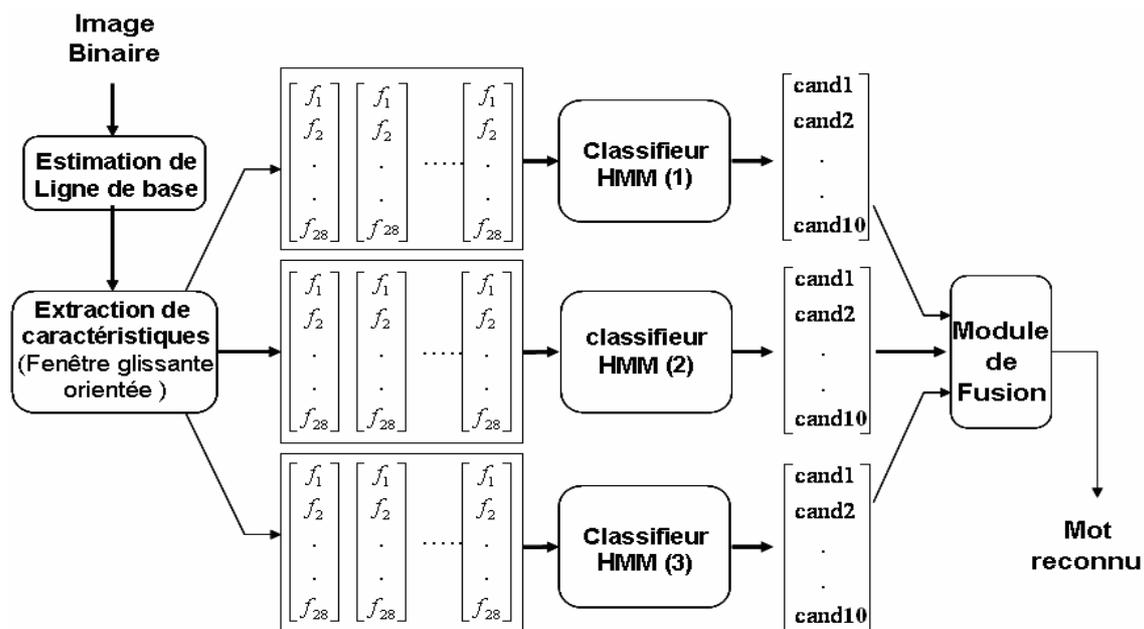


FIG. 1 - Diagramme du système proposé

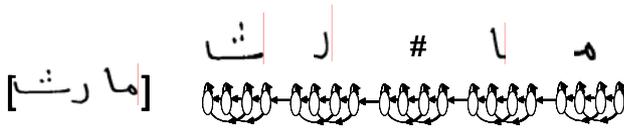


FIG. 2 - Topologie d'un modèle HMM de mot

Le système générique HCM [MOK 02] réalise l'apprentissage et la reconnaissance des mots et caractères. L'apprentissage des modèles de caractères se fait à partir des images de mots et de leur transcription [ELH 06]

La reconnaissance utilise l'algorithme de Viterbi. La séquence d'observations issue du mot à reconnaître est confrontée au réseau des mots du lexique composé de modèles de caractères. La séquence de caractères correspondant à la vraisemblance maximale identifie le mot ou la phrase reconnue.

3 Système à fenêtres orientées

La figure 1 illustre les différentes étapes du système combiné à fenêtres orientées. L'idée principale consiste à intégrer une information sur l'inclinaison possible de l'écriture en divisant l'image binaire d'un mot en fenêtres verticales et inclinées. Une solution serait de rechercher l'inclinaison globale des traits pour chaque mot, de redresser les traits inclinés et de fournir l'image corrigée au système de référence (cf. Section 4). Pour éviter cette recherche d'inclinaison mot à mot, nous inclinons les fenêtres des deux côtés de l'axe vertical de manière systématique. Une de ces fenêtres au moins s'approchera de l'inclinaison réelle du mot (sans inclinaison, orienté à gauche ou orienté droite). Les fenêtres inclinées peuvent d'autre part intercepter les points et marques diacritiques décalés par rapport à leur position idéale.

Au système de référence sont associés deux autres classifieurs HMMs utilisant des fenêtres inclinées, symétriquement ou non. La méthode présentée dans ce travail consiste à extraire des vecteurs de primitives sur des fenêtres inclinées (orientées) dans le but de remédier aux problèmes de points diacritiques et d'inclinaison.

La première étape du système consiste à transformer l'image binaire d'un mot en trois séquences de vecteurs de caractéristiques (primitives). Et ceci en divisant l'image en fenêtres verticales avec trois angles d'inclinaisons successives $(-\alpha \ 0 \ +\alpha)$ selon le principe de la fenêtre glissante de droite à gauche (Fig. 3). Ensuite, un vecteur de primitives est calculé pour chaque fenêtre.

Les fenêtres avec le même angle d'inclinaison se chevauchent. La longueur de chevauchement est un paramètre du système et qui peut varier de 0 pixels à *largeur-de-fenêtre* - 1 pixels. En plus, les fenêtres qui ont le même indice i ($i=0, \dots, nb\text{-de-fenêtre} - 1$) et avec des inclinaisons différentes se croisent sur la ligne d'écriture de base (Fig. 3). Les mêmes caractéristiques

que celles utilisées dans le système de référence sont extraites sur chaque fenêtre et chaque inclinaison.

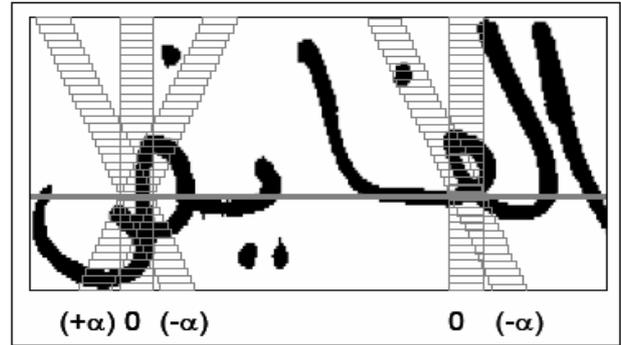


FIG. 3 - Croisement des fenêtres inclinées sur la ligne d'écriture (ici symétriquement)

Le résultat du module d'extraction de vecteurs de caractéristiques est trois suites de vecteurs de primitives. Chaque suite correspond à une séquence de fenêtres inclinées ou non et est destinée à être traitée par un de trois classifieurs HMM.

Les systèmes HMMs utilisés sont de même type que celui utilisé dans le système de référence. Ainsi, chaque classifieur HMM produit une liste de 10 meilleurs candidats (top 10) qui ont les plus grandes vraisemblances (Fig. 1).

En deuxième phase du système et en post-traitement, un algorithme permet de combiner les classifieurs en fusionnant les listes de candidats proposées par les classifieurs HMM. La fusion se fait en accumulant les log- vraisemblances des solutions dans les trois listes. Les vraisemblances accumulées servent à réarranger les candidats, en déplaçant vers le haut les candidats les plus appropriés

4 Expériences et résultats

Les résultats préliminaires sur la combinaison des classifieurs ont été présentés dans [ELH 07]. Dans cette section on présente les résultats complets des expériences sur la variation des valeurs des angles d'inclinaison, la comparaison des combinaisons symétriques et asymétriques des classifieurs, ainsi que la comparaison du système de combinaison avec un système de référence comprenant une étape de correction d'inclinaison en pré-traitement.

Nous testons tout d'abord les classifieurs individuels (non combinés) pour différentes valeurs d'angles. La Figure 4 donne les performances de ces classifieurs HMMs individuels pour des orientations de fenêtres entre -45° à $+45^\circ$. Pendant les expériences on considère les 946 classes des noms des villes de la base IFN-ENIT. Les sous ensembles *a*, *b*, et *c* de cette base ont été utilisés pour l'entraînement des modèles des caractères. Tandis que le sous ensemble *d* a été utilisé pour le test. Les fenêtres de faible inclinaison ($0^\circ, 5^\circ, 10^\circ, 15^\circ$) à gauche ou à droite donnent les meilleurs résultats pour les

systèmes individuels. Cela signifie que l'observation des images des mots avec des fenêtres proches de la verticale est meilleure qu'avec des fenêtres fortement orientées.

Puis, nous testons la combinaison des classifieurs pour des angles symétriques (Fig. 5, courbe Top1 Symétrique). Ces expériences montrent que la combinaison du classifieur de référence avec les

classifieurs orientés est efficace pour des angles entre 10° et 25° . Le tableau 1 donne les performances des classifieurs individuels et du classifieur combiné pour des fenêtres orientées de 20° de part et d'autre de la verticale. L'amélioration est significative par rapport au classifieur orienté verticalement.

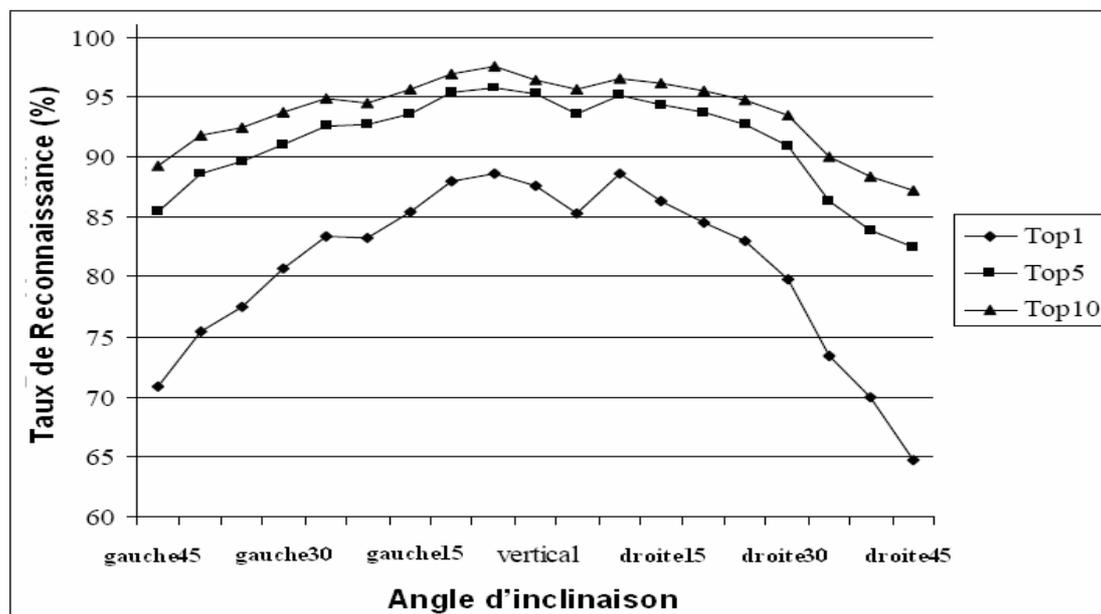


FIG. 4 - Taux de reconnaissance des classifieurs individuels avec des angles d'orientation différents

Classifieur	Taux de reconnaissance (%)		
	Top1	Top2	Top3
HMM1 (0°)	87.60	91.42	93.76
HMM2 ($\alpha = +20^\circ$)	83.31	88.42	90.82
HMM3 ($\alpha = -20^\circ$)	84.59	89.28	91.76
Combinaison par somme des log-vraisemblances	90.61	94.89	95.87

TAB. 1 – Taux de reconnaissance: système de référence (HMM1) et système combiné par sommation des scores. Sous-ensemble de test: d, nombre de classes:946

Des expériences ont ensuite été réalisées sur la variation des angles des classifieurs à combiner. La Figure 5 donne les résultats du système combiné (somme des log-vraisemblances) quand on introduit une dissymétrie dans l'orientation de la fenêtre inclinée à gauche avec celle inclinée à droite. Les performances sont meilleures que pour n'importe quel HMM individuel et le système combiné est opérationnel pour une large gamme d'angles. Dans la figure 5, et pour chaque angle d'inclinaison x , on spécifie une étiquette '*Comb x* ' sur l'axe des abscisses. L'angle x est l'angle de combinaison de base pour une combinaison à fenêtres symétriques. Cet angle de base x est ensuite augmenté d'une valeur dx pour une des fenêtres, celle de droite (courbe droite+ dx) ou celle de gauche (courbe gauche+ dx). La valeur d'angle de décalage $dx \in \{0, 5, 10, 15\}$. $dx=0$ correspond à une combinaison symétrique (courbe Top 1 Symétrique). Pour $x=10^\circ$,

le point de la courbe droite+5 ($dx=5$ pour la fenêtre de droite) correspond à une combinaison asymétrique des classifieurs d'angles $(-10, 0, 15)$ et le point de la courbe gauche+5 ($dx=5$ pour la fenêtre de gauche) à une combinaison asymétrique $(-15, 0, 10)$. Pour la courbe (gauche+5), les performances sont bonnes pour les systèmes utilisant les triplets d'angles $(-20^\circ, 0^\circ, 15^\circ)$, $(-25^\circ, 0^\circ, 20^\circ)$, $(-30^\circ, 0^\circ, 25^\circ)$, $(-35^\circ, 0^\circ, 30^\circ)$, notons que les angles sont donnés dans l'ordre suivant : orientation gauche, orientation verticale, orientation droite.

La dernière étape de nos d'expériences consiste à comparer l'approche proposée avec un système de référence comprenant une étape de correction d'inclinaison en pré-traitement. On a utilisé la méthode de correction d'inclinaison KSC proposée dans [KIM 93]. Cette méthode est basée sur la direction des contours des points d'écriture pour estimer l'inclinaison moyenne d'un mot. Ce système est comparé aux classifieurs individuels et aux combinaisons de ces classifieurs (somme des log-vraisemblances) pour différentes valeurs d'inclinaison des fenêtres.

Dans la figure 6, on présente les taux de reconnaissance obtenus pour différents angles d'inclinaison (spécifiés sur l'axe des abscisses) par les classifieurs individuels (les courbes : Top1, Top5, et Top10), les systèmes combinés à fenêtres symétriques (les courbes : Combinaison Top1, Combinaison Top5, et Combinaison Top10), et par les classifieurs avec une étape de correction d'inclinaison en pré-traitement (les

courbes : Pre-traitement Top1, Pre-traitement Top5, et Pre-traitement Top10). Les résultats présentés en cette Figure montrent que le système utilisant la correction d'inclinaison en pré-traitement est meilleur que le système de référence à fenêtres verticales. D'autre part, le système combiné de classifieurs (avec

fenêtres symétriques) est meilleur que le système à fenêtres verticales et correction d'inclinaison, pour les angles de fenêtre entre 10° et 25°. Cela démontre l'avantage du système de combinaison proposé.

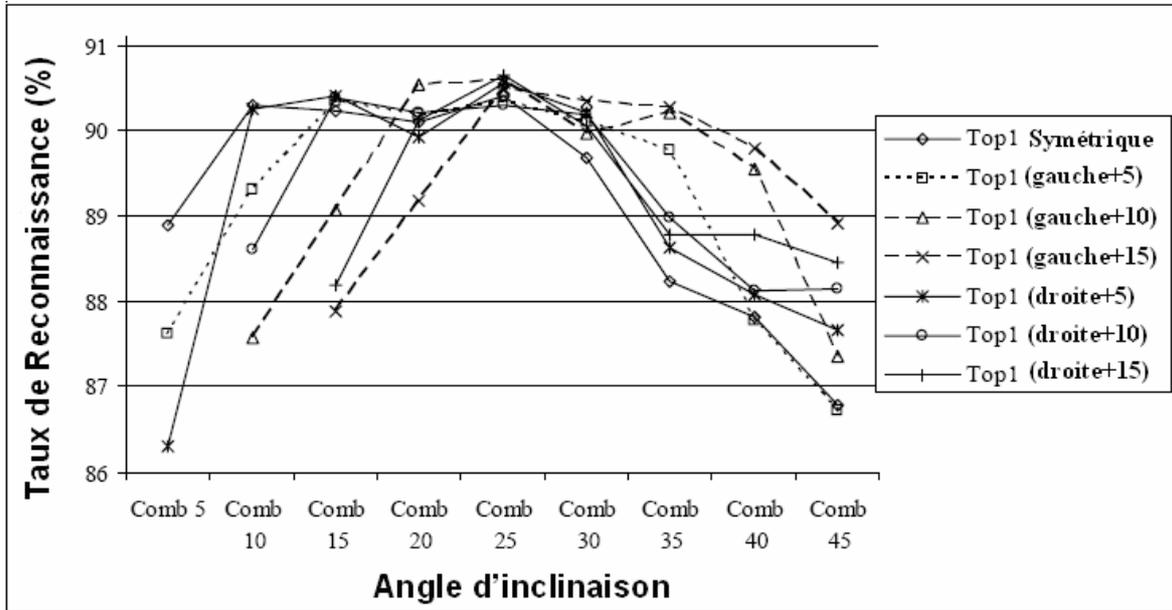


Fig. 5 – Taux de reconnaissance obtenus par combinaison des trois classifieurs pour différents angles d'inclinaison des fenêtres

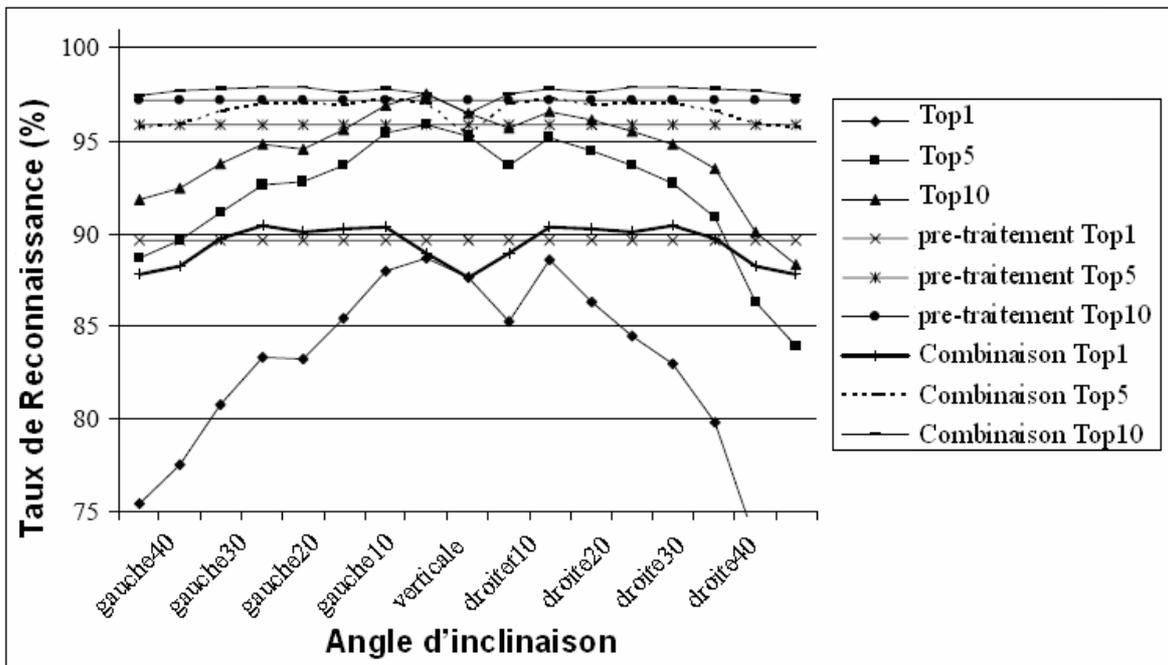


Fig. 6 – Taux de reconnaissance obtenus pour différents angles d'inclinaison, par les classifieurs individuels, le système combiné (fenêtres symétriques), et par un classifieur utilisant les fenêtres verticales et une étape de correction d'inclinaison en pré-traitement

5 Conclusions et Perspectives

Dans cet article nous avons illustré l'avantage d'extraire des caractéristiques sur des fenêtres verticales et orientées pour la reconnaissance hors ligne des mots manuscrits cursifs arabes. La méthode proposée est composée d'une étape de reconnaissance et d'une étape de combinaison en post-traitement. Cette méthode prend en considération les inclinaisons ainsi que les positions erronées des points et des marques diacritiques possibles dans l'écriture qui sont des sources majeures d'erreurs pour la reconnaissance des mots.

L'étape de reconnaissance est de type HMM et est basée sur 3 classifieurs qui utilisent des fenêtres glissantes de différentes orientations. L'étape de combinaison fusionne les sorties des HMMs par la somme des log-vraisemblances. La combinaison a été examinée pour différentes valeurs d'inclinaison des fenêtres, et suivant des orientations symétriques et non symétriques. Les résultats obtenus sur la base de référence IFN/ENIT montrent que la combinaison des classifieurs, et pour des valeurs spécifique des angles d'orientation des fenêtres, est meilleure que le système simple incluant une étape de correction d'inclinaison d'écriture.

Les perspectives futures comprennent la proposition d'une règle de rejet pour détecter les mots ambigus, ainsi que la combinaison de notre système à d'autres systèmes utilisant des approches similaires avec autres types de caractéristiques. Une perspective majeure consiste à optimiser la stratégie de combinaison en post-traitement.

Remerciement

Ce travail a été partiellement supporté par le programme de recherche SARIMA.

(<http://www.sarima.org> ou <http://www.irisa.fr/sarima>)

Références

- [BEN 03] Ben Amara N., Bouslama F. (2003), Classification of Arabic script using multiple sources of information: state of the art and perspective, *IJDAR*, Vol. 5, pp. 195-212.
- [BEN 00] Ben Amara N., Belaïd A., Ellouze N., Utilisation des modèles Markoviens en reconnaissance de l'écriture arabe état de l'art, *Actes de CIFED'00*, pp. 181-191, Lyon, France, (2000).
- [ELH 07] El-Hajj R., Mokbel C., Likforman-Sulem L., Combination of HMM-based classifiers for the recognition of Arabic Handwritten Words, *Actes d'ICDAR'07*, pp. 959-963, Curitiba – Brazil, (2007).
- [ELH 06] El-Hajj R., Mokbel C., Likforman-Sulem L., Reconnaissance de l'écriture Arabe cursive: combinaison de classifieurs HMMs à fenêtres orientées, *Actes de CIFED'06*, pp. 271– 276, Fribourg – Suisse, (2006).
- [ELH 05] EL-Hajj R., Likforman-Sulem L., Mokbel C., Arabic handwriting recognition using baseline dependent features and Hidden Markov Modeling, *Proc. of ICDAR 05*, pp. 893-897, Seoul (2005).
- [KHO 03] Khorsheed M.S. (2003), Recognizing Arabic manuscripts using a single hidden Markov model, *Pattern Recognition Letters*, Vol. 24, pp. 2235-2242.
- [KIM 93] Kimura F., Shridhar M., Chen Z., Improvements of a lexicon directed algorithm for recognition of unconstrained handwritten words, *Actes d' ICDAR'93*, pp.18-22, Tsukuba, Japan, (1993).
- [LOR 05] Lorigo L.M., Govindaraju V., Segmentation and pre-recognition of Arabic handwriting, *Actes d'ICDAR'05*, pp. 605 – 609, Seoul, (2005)
- [MAR 05] Märgner V., Pechwitz M., El-Abed H., Arabic Handwriting Word Recognition Competition, *Actes d'ICDAR'05*, pp. 70 - 74, Seoul, (2005).
- [MOH 96] Mohamed M.A., Gader P.D. (1996), Handwritten word recognition using segmentation-based hidden Markov modeling and segmentation-based dynamic programming techniques, *IEEE PAMI*, Vol. 18, pp. 548–554.
- [MOK 02] Mokbel C., Abi Akl H., Greige H., Automatic speech recognition of Arabic digits over the telephone network, *Proc. of Research Trends in Science and Technology RSTS'02*, Beyrouth (2002).
- [PEC 03] Pechwitz M., Märgner V. (2003), HMM Based approach for handwritten Arabic Word Recognition Using the IFN/ENIT– DataBase, *Actes d'ICDAR'03*, pp. 890-894, Edinburgh (2003).
- [PEC 02] Pechwitz M., Maddouri S., Märgner V., Ellouze N. (2002), IFN/ENIT–DataBase for Handwritten Arabic words, *Actes de CIFED'02*, pp 129-136, Hammamet, Tunisia (2002).
- [SCH 96] Schwartz R., LaPre C., Makhoul J., Raphael C., Zhao Y. (2006), Language-Independent OCR Using a Continuous Speech Recognition System. *Actes d'ICPR'96*, pp. 99-103, Vienne, Autriche, (1996).