



HAL
open science

Reconnaissance de l'écriture arabe cursive : combinaison de classifieurs MMCs à fenêtres orientées

Ramy Al-Hajj, Chafic Mokbel, Laurence Likforman-Sulem

► To cite this version:

Ramy Al-Hajj, Chafic Mokbel, Laurence Likforman-Sulem. Reconnaissance de l'écriture arabe cursive : combinaison de classifieurs MMCs à fenêtres orientées. Sep 2006, pp.271-276. hal-00111156

HAL Id: hal-00111156

<https://hal.science/hal-00111156>

Submitted on 3 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.

Reconnaissance de l'écriture arabe cursive : combinaison de classifieurs MMCs à fenêtres orientées

Ramy Al-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 likforman@tsi.enst.fr Chafic.Mokbel@balamand.edu.lb

Résumé : *L'écriture arabe est naturellement cursive, difficile à segmenter et présente une grande variabilité. Nous présentons tout d'abord un système de référence pour la reconnaissance hors-ligne de l'écriture cursive basé sur les modèles de Markov cachés (MMC), de type analytique et sans segmentation. Les vecteurs de caractéristiques sont liés aux densités et aux configurations locales de pixels. Ces vecteurs sont extraits sur les images binaires de mots, de droite à gauche, par des fenêtres glissantes. Le système générique HCM, à l'état de l'art, apprend les modèles de caractères et de mots à partir de leur image et de leur transcription. Puis nous proposons une méthode avancée qui prend en considération les inclinaisons possibles dans l'écriture. Cette méthode est composée de deux étapes, une étape de reconnaissance et une étape de combinaison en post-traitement. L'étape de reconnaissance est basée sur l'extraction des vecteurs de caractéristiques toujours selon la technique des fenêtres glissantes mais avec des inclinaisons différentes permettant d'obtenir dans la phase de classification, et pour chaque inclinaison, un ensemble de mots candidats comme résultat. L'étape de combinaison permet de fusionner en post-traitement les résultats obtenus pour produire le candidat le plus approprié. Le système a montré de bonnes performances sur la base IFN/ENIT de noms de ville Tunisiens.*

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

1 Introduction

La reconnaissance des textes cursifs reste toujours un problème ouvert aussi bien dans sa forme imprimée que manuscrite. Ceci à cause des difficultés auxquelles sont confrontés les chercheurs et les développeurs, telles que la variabilité de la forme, du style, et de l'inclinaison de l'écriture manuscrite.

L'écriture manuscrite arabe est naturellement cursive, difficile à segmenter, et présente une grande variabilité. Les ligatures entre les caractères ainsi que les espaces inter et intra-mots ont des longueurs différentes. 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. De plus, les caractères d'un mot, ou d'un pseudo mot, sont ligaturés horizontalement sur la ligne d'écriture dite ligne de base, et des ligatures verticales de deux ou trois caractères sont possibles.

A cause de la nature cursive et connectée de l'écriture Arabe imprimée et manuscrite, plusieurs systèmes basés sur les MMCs ont été développés pour la reconnaissance [BEN 00][KHO 03][MAK 98][MIL 97][PEC 03]. L'avantage des systèmes MMCs est qu'ils offrent une approche segmentation-reconnaissance conjointe et que, même dans le cas des approches analytiques, les MMCs permettent d'éviter la pré-segmentation des mots en caractères dans la phase d'entraînement, et dans la phase de classification. De plus, les systèmes MMCs permettent, en modélisant stochastiquement les séquences d'observations de longueurs variables, de traiter les distorsions non linéaires.

Dans cet article nous décrivons tout d'abord notre système de référence pour la reconnaissance hors-ligne de l'écriture cursive basé sur les modèles de Markov cachés, de type analytique et sans segmentation. Les vecteurs de caractéristiques sont extraits de droite à gauche sur les images binaires de mots par des fenêtres verticales glissantes intercalées. La reconnaissance est basée sur un ensemble de caractéristiques niveau pixels, robustes et indépendantes du langage d'écriture. Des paramètres tels que les positions des lignes d'écriture de base sont utilisées pour dériver un ensemble de primitives dépendantes de ces lignes. Ainsi, les variabilités d'écriture dans les deux parties haute et basse d'un mot sont bien prises en compte.

Est présentée ensuite une méthode de reconnaissance qui prend en considération les inclinaisons possibles dans l'écriture. Il s'agit d'un système à deux étapes : la première étape concerne la reconnaissance des mots par trois classifieurs issus du système de référence, la deuxième étape dite de post-traitement, est une étape de combinaison par fusion des scores des mots-candidats

proposés lors de la première étape. L'extraction des primitives est faite en utilisant la technique des fenêtres glissantes mais avec des inclinaisons différentes permettant d'obtenir dans la phase de classification, et pour chaque inclinaison α , un ensemble de mots candidats comme résultats. L'étape de post-traitement permet de fusionner les résultats obtenus pour produire le candidat le plus approprié.

Les deux systèmes, système de référence et système avancé, sont basés sur une approche analytique. Un modèle spécifique est utilisé pour chaque caractère. Et les modèles des mots sont construits par concaténation des modèles des caractères constituants.

Cet article insiste sur l'efficacité du choix des caractéristiques (primitives) utilisées sur la performance du système de reconnaissance. Ces caractéristiques sont passées au système générique HCM [MOK 02] pour apprendre, et ensuite tester, les modèles de caractères et de mots à partir de leur image et de leur transcription. De plus, cet article signale l'importance de l'information sur l'inclinaison qui est intégrée en post-traitement. La Figure 1 illustre les différentes étapes du système de référence.

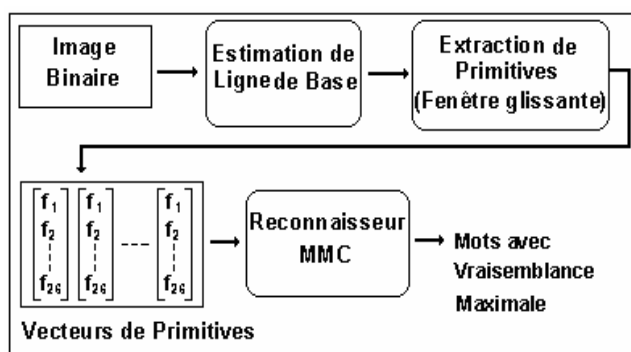


Fig. 1 - Diagramme du système de référence

Dans la section 2, nous présentons le système de référence et les caractéristiques extraites en utilisant la technique des fenêtres verticales glissantes. La troisième section détaille le module MMC de classification utilisé dans notre système. Dans la section 4 nous présentons le système de reconnaissance avancé intégrant l'information sur l'orientation dans l'écriture. Les performances des deux systèmes testés sur la base de données de référence IFN/ENIT de noms de ville Tunisiens [PEC 02] sont présentées dans la section 5. Finalement, des conclusions ainsi que des perspectives futures sont présentées dans la section 6.

2 Système de référence à base de modèles de Markov cachés

Une première version de ce système est décrite dans [ELH 05]. Deux caractéristiques ou paramètres ont été ajoutés aux vecteurs de paramètres dans le système de référence présenté ici. Ce système a obtenu de très 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 [ELH 05] (cf. Figure 2). Cette étape permet de séparer l'image du mot en 3 zones : une zone médiane qui correspond à la zone centrale de l'écriture, une zone supérieure où se trouvent les hampes des caractères et les points diacritiques hauts, une zone inférieure où se trouvent les jambages des caractères et les points diacritiques bas. 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]. Ces fenêtres sont de largeur $k=8$ pixels et sont successivement décalées de δ pixels (δ est un paramètre qui prend ses valeurs entre 1 et $k-1$). Dans chaque fenêtre est extrait un ensemble de 26 caractéristiques. Celles-ci sont représentatives des densités et des configurations locales des pixels d'écriture. Pour cela, chaque fenêtre est divisée en un nombre de cellules fixe. 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.

2.1 Détection des lignes de base

Généralement, l'extraction de la ligne de base est utilisée dans les phases ultérieures de normalisation et de segmentation des mots. Dans notre système nous utilisons les positions des lignes d'écriture: haute et basse, pour dériver un ensemble de caractéristiques dépendantes de ces lignes. La Figure 2 présente deux exemples sur des lignes de base détectées par notre système.

Notre approche est basée sur l'algorithme décrit dans [4] avec quelques modifications. En effet, cette approche utilise le profil de projection vertical des densités des pixels le long de l'axe horizontal. Premièrement, la ligne de base basse correspondante au maximum de la courbe de profil de projection, est cherchée. Ceci est justifié dans le cas de la langue écrite arabe où la plupart des lettres ont beaucoup de pixels sur la ligne de base basse.

Après, l'algorithme parcourt l'image du mot de haut au bas pour trouver la ligne de base supérieure (haute) correspondante à la première ligne avec une valeur de projection plus grande ou égale à la densité de rang moyenne.

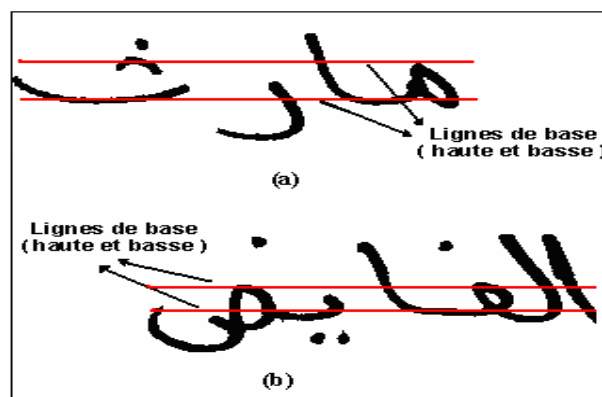


Fig. 2 - Estimation de la position des lignes de base.

2.2 Caractéristiques de densités

Ces caractéristiques sont au nombre de 16. Nous appelons L et U, les positions (ordonnée y) de la ligne de base basse et haute respectivement. Soit H, la hauteur en pixels de la fenêtre glissante (hauteur de la boîte englobante du mot traité). La fenêtre étant divisée verticalement en n_c cellules, soit $n(i)$: le nombre de pixels d'écriture (pixels noirs) dans la cellule i. Soit $b(i)$: l'intensité de la cellule i :

$$b(i)=0 \text{ si } n(i)=0, b(i)=1 \text{ sinon.}$$

Les caractéristiques de densités sont les suivantes:

f_1 : densité des pixels noirs dans la fenêtre

f_2 : nombre de transitions Noir/Blanc entre cellules

$$f_2 = \sum_{i=2}^{n_c} |b(i) - b(i-1)|$$

f_3 : différence de position entre les centres de gravité g des pixels d'écriture dans deux fenêtres consécutives :

$$f_3 = g(t) - g(t-1)$$

f_4 à f_{11} : sont les densités de pixels d'écriture dans chaque colonne de la fenêtre.

Les caractéristiques suivantes dépendent de la position estimée des lignes de base.

f_{12} : position verticale normalisée du centre de gravité des pixels d'écriture, par rapport à la ligne de base basse.

$$f_{12} = \frac{g - L}{H}$$

f_{13} - f_{14} : densités des pixels d'écriture au dessus et au dessous de la ligne de base basse

f_{15} : nombre de transitions Noir/Blanc entre les cellules situées au dessus de la ligne de base basse

f_{16} : zone à laquelle appartient le centre de gravité de l'écriture dans la fenêtre (zone supérieure $f_{16}=1$, zone médiane $f_{16}=2$, zone inférieure $f_{16}=3$)

2.3 Caractéristiques de configurations locales

Les caractéristiques f_{17} à f_{21} sont liées aux configurations locales des pixels d'écriture. Dans chaque fenêtre d'analyse, on cherche le nombre de pixels du fond (blancs) qui appartiennent à l'une des 5 configurations de la Figure 3. Le nombre de pixels dans chaque configuration est ensuite normalisé par le nombre de pixels dans la fenêtre.

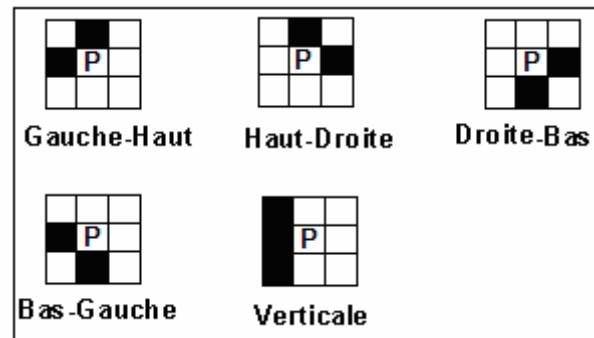


Fig. 3 - Les 5 types de configurations locales autour d'un pixel de fond P

Les 5 caractéristiques suivantes dépendent des positions des lignes de base. f_{22} à f_{26} sont aussi le nombre de pixels de fond dans les 5 configurations citées précédemment mais seulement pour les pixels se situant dans la zone médiane de l'écriture, entre les deux lignes de base.

L'ensemble des 26 caractéristiques extraites comporte 10 caractéristiques qui dépendent de la position des lignes de base, et 16 qui n'en dépendent pas. Par ailleurs, cet ensemble est indépendant du langage utilisé. Ces caractéristiques peuvent être utilisées pour tout type d'écriture cursive comportant des hampes et des jambages comme l'écriture Arabe et l'écriture Latine.

3 Modélisation markovienne analytique

Le système de reconnaissance modélise les mots et les caractères sous la forme de modèles de Markov cachés (MMC). 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 4. Les densités de probabilité des observations dans chaque état sont modélisées par un mélange de 3 gaussiennes. Au total, 159 modèles de caractères sont construits pour l'écriture arabe pour tenir compte du fait que certains caractères ont des formes différentes suivant leur position dans le mot (au début, au milieu ou à la fin), et de la présence de marques additionnelles telles que le "shadda" (ّ).

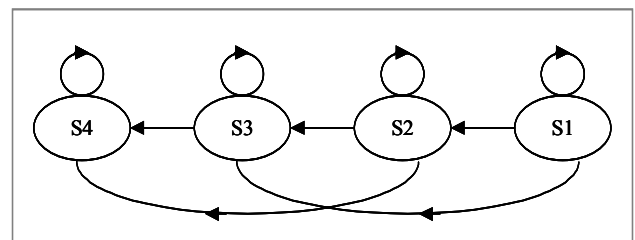


Fig. 4 - Topologie d'un modèle MMC de caractère

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.

L'approche est donc sans segmentation car les modèles de caractères sont appris par apprentissage croisé [Choisy 2003] sans avoir à segmenter au préalable les mots en caractères.

L'apprentissage utilise l'algorithme itératif EM de la façon suivante : les observations sont associées aux états les plus probables pendant la phase d'espérance (E), les paramètres des modèles sont re-estimés pendant la phase de maximisation (M). Lors de l'initialisation, les observations sont affectées dans les états en les segmentant linéairement puis les premiers paramètres sont estimés à partir de cette affectation.

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.

4 Système avancé

Après avoir mis en œuvre un système de reconnaissance hors ligne des mots manuscrits cursifs, basé sur une approche MMC analytique sans segmentation, et qui donne des résultats à l'état de l'art [MAR 05][ELH 05], nous proposons un système de reconnaissance plus avancé. L'idée principale consiste à intégrer une information sur l'inclinaison possible dans l'écriture en divisant l'image binaire d'un mot en fenêtres verticales et inclinées.

En fait, parmi les difficultés rencontrées lors de la reconnaissance de l'écriture Arabe cursive, on peut citer le problème des positions erronées des points et des marques diacritiques au dessus et au dessous des caractères, ainsi que le problème d'inclinaison de l'écriture. Il est à noter que ces problèmes se trouvent accentués dans le cas de manuscrits à cause de la diversité des formes et de la variabilité des scripts. La méthode présentée dans ce travail consiste à extraire des vecteurs de primitives sur des fenêtres inclinées (orientées) dans le but de remédier à ces deux types de problèmes.

4.1 Description du système

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 (Figure 5) (Couramment, la valeur de α est un paramètre du système). Ensuite, un vecteur de primitives est calculé pour chaque fenêtre.

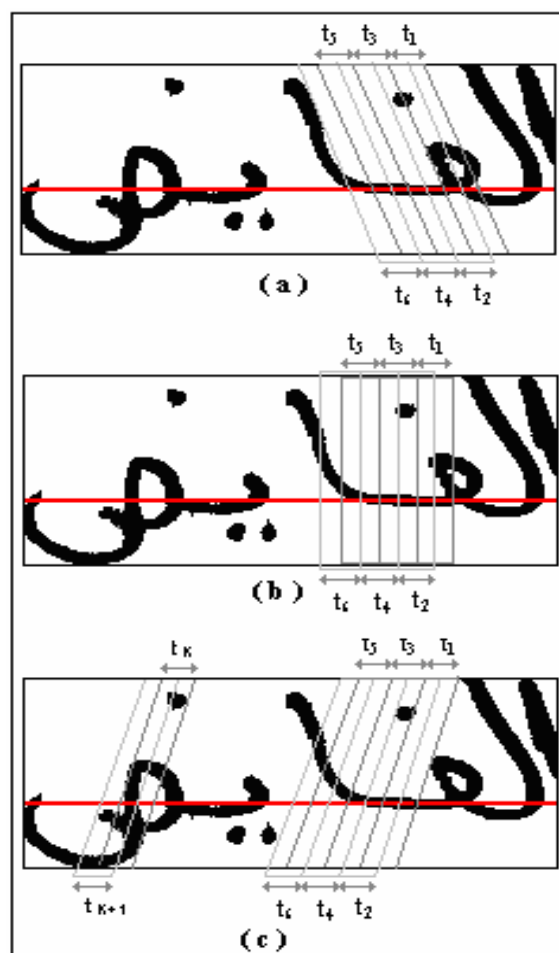


Fig. 5 - Extraction de vecteurs de primitives avec des fenêtres inclinées et chevauchantes et trois types d'inclinaisons : vers la gauche (a), verticale (b), vers la droite (c)

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 (Figure 6).

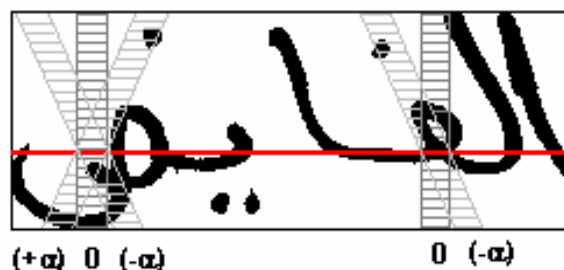


FIG. 6 - Croisement des fenêtres inclinées sur la ligne d'écriture

La hauteur d'une fenêtre varie selon la dimension de chaque image. Chaque fenêtre est divisée en cellules. Le nombre de cellules est toujours fixé à 21 cellules pour chaque fenêtre. Et les mêmes caractéristiques que celles

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

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 classificateurs MMC. Les systèmes MMCs utilisés sont de même type que celui utilisé dans le système de référence. Ainsi, chaque classificateur MMC produit une liste de 4 meilleurs candidats (top 4) qui ont les plus grandes vraisemblances (Figure 7).

En deuxième phase du système et en post-traitement, un algorithme permet de combiner les classificateurs en fusionnant les listes de candidats proposées par les classificateurs MMC. 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. La Figure 7 montre l'architecture générale du système proposé.

Reconnaisseur	Taux de Reconnaissance (%)		
	Top1	Top2	Top3
Classifieur-1 (+ α°)	82.19	86.57	88.98
Classifieur-2 (0°) (système de référence)	84.03	88.47	90.74
Classifieur-3 (- α°)	82.47	87.19	89.56
Système proposé	85.02	91.29	93.14

TAB. 1 – Résultats comparatifs pour le système avancé et le système de référence

du taux d'erreur) de la performance en reconnaissance. Ceci démontre que l'information apportée en inclinant les fenêtres est complémentaire.

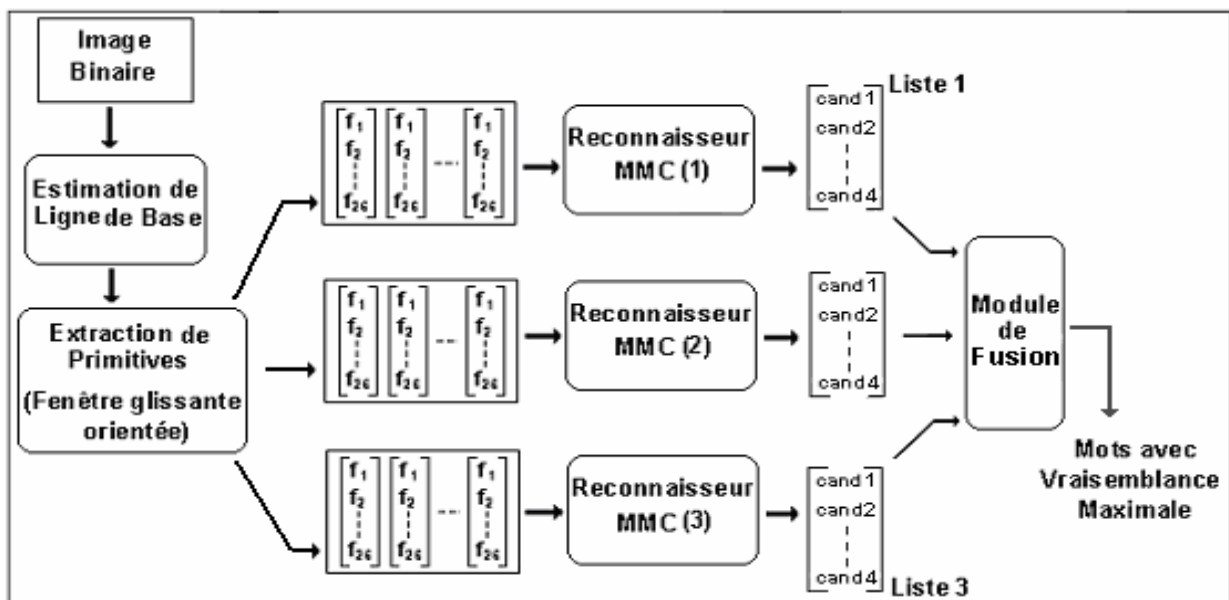


FIG. 7 - Architecture générale du système avancé

5 Expériences et résultats

Pour évaluer la performance de la méthode proposée, des expériences ont été réalisées sur un sous ensemble de la base de référence IFN/ENIT. Ainsi, les expériences d'évaluation ont été dérivées sur un ensemble de 16128 images pour l'apprentissage, et un ensemble de 5352 images pour le test, et avec un dictionnaire de taille 459 pour les noms de villes. Le tableau 1 présente les résultats du système proposé par rapport aux résultats obtenus par le système de référence. La valeur du taux de reconnaissance (resp. taux d'erreur), pour le Top1, croît de 84.03% à 85.02% (resp. décroît de 15.97% à 14.98%). Ce qui montre que la méthode proposée produit une amélioration notable (1.0 % en valeur absolue, ou une réduction de 6.27 %

6 Conclusions et Perspectives

Dans cet article un système de référence pour la reconnaissance hors-ligne de l'écriture cursive est présenté. Le système est de type analytique, sans segmentation et basé sur les modèles de Markov cachés, et l'extraction d'un ensemble de caractéristiques robustes et indépendantes du langage d'écriture. Ce système a obtenu de bons résultats à la compétition organisée sur la base de données de référence IFN/ENIT à ICDAR 05 [MAR 05]. Un système avancé qui prend en considération les inclinaisons possibles dans l'écriture manuscrite est ensuite introduit. Ce système opère en deux étapes, une étape de reconnaissance basée sur l'extraction des caractéristiques avec des fenêtres glissantes orientées, et une étape de combinaison basée sur la fusion des

résultats de classifieurs MMCs associés à différentes inclinaisons. La performance de reconnaissance du système avancé proposé a été évaluée

sur un sous-ensemble de la base de référence IFN/ENIT. Les résultats obtenus montrent une amélioration significative démontrant que l'introduction d'information sur l'inclinaison de l'écriture ne peut qu'améliorer la reconnaissance.

Plusieurs tâches restent encore à achever, telles que la proposition et l'examen de nouvelles primitives, et l'adaptation des lignes de base pour chaque composant d'un mot composé. Une perspective majeure consiste à améliorer la phase de fusion en introduisant un réseau de neurones multicouches (MLP) capable de fusionner et de vérifier les résultats obtenus par les modules de classification MMCs.

Remerciement

Ce travail a été partiellement financé par le programme de recherche SARIMA (<http://www.sarima.org> ou <http://www.irisa.fr/sarima>).

Références

- [AMI 98] Amin A. (1998), Off-line Arabic character recognition: the state of the art, *Pattern Recognition*, Vol. 31, 5, pp 517-530.
- [ARI 02] Arica N., Yarman-Vural F.T, (2002) Optical character recognition for cursive handwriting, *IEEE PAMI*, Vol. 24, 6, pp 801 – 813.
- [BEN 00] Ben Amara N., Belaïd A., Ellouze N. (2000), Utilisation des modèles markoviens en reconnaissance de l'écriture arabe état de l'art. *In Colloque International Francophone sur l'Ecrite et le Document (CIFED'00)*, Lyon, France.
- [BLU 02] Blumenstein M., Cheng C. K., Liu X. Y. (2002), New Preprocessing techniques for Handwritten Word Recognition, *Proc. of the 2nd IASTED Conf. on Visualization, Imaging and Image Processing*, pp. 480-484.
- [KHO 03] Khorsheed M.S. (2003), Recognizing Arabic manuscripts using a single hidden Markov model, *Pattern Recognition Letters*, 24, pp. 2235-2242.
- [MAK 98] Makhoul J., Schwartz R., Lapre C., Bazzi I. (1998), A script independent methodology for optical character recognition, *Pattern Recognition*, vol. 31, no 9, pp. 1285-1294.
- [MIL 97] Miled H., Olivier C., Cheriet M., Lecourtier Y. (1997), Coupling observation/letter for a Markovian modelization applied to the recognition of Arabic handwriting, *ICDAR 97*, Ulm, pp 580 - 583 .
- [MOK 02] Mokbel C., Abi Akl H., Greige H. (2002), Automatic speech Recognition of Arabic digits over the telephone network, *Proc. of Research Trends in Science and Technology RSTS'02*, Beyrouth.
- [PEC 02] Pechwitz M., Maddouri S., Märgner V., Ellouze N. (2002), IFN/ENIT–DataBase for Handwritten Arabic words, *CIFED'02*, Hammamet, Tunisia, pp 129-136.
- [PEC 03] Pechwitz M., Märgner V. (2003), HMM Based approach for handwritten Arabic Word Recognition Using the IFN/ENIT– DataBase, *ICDAR'03*, Edinburgh, pp. 890-894.
- [VIA 99] C. Viard-Gaudin, PM Lallican, S. Knerr, P. Binter (1999), The IRESTE on/off (IRONOFF) dual handwriting database, *ICDAR 99*, pp.455-458.
- [MAR 99] U. Marti and H. Bunke (1999) A full English sentence database for off-line handwriting recognition, *ICDAR'99*, pp 705 - 708.
- [CHO 03] Choisy C., Modélisation analytique de l'écriture manuscrite par une approche optimale sans segmentation basée sur des champs de Markov, Thèse de Doctorat, Université de Nancy 2, 2003
- [MAR 05] Märgner V., M. Pechwitz, H. El-Abed, Arabic Handwriting Word Recognition Competition, Actes de ICDAR 05, Seoul, Aout 2005, p. 70-74.
- [ELH 05] R. El-Hajj, L. Likforman-Sulem, C. Mokbel, Arabic handwriting recognition using baseline dependent features and Hidden Markov Modeling, *ICDAR 05*, Seoul, Corée du Sud, Août 2005.
- [SCH 96] R. Schwartz, C. LaPre, J. Makhoul, C. Raphael, Y. Zhao. Language-Independent OCR Using a Continuous Speech Recognition System. *ICPR'96*. Volume 3, p.99.