

Reconnaissance automatique de caractères et de textes amazighes : état des lieux et perspectives

Ali Rachidi

ENCG et Laboratoire IRF-SIC, Université Ibn Zohr, Agadir, Maroc

The design and implementation of systems OCR Amazigh character is very crucial for the promotion and development of the Amazigh language. To date, there is the lack of this type of system. Therefore, the automatic character recognition and text Amazigh has experienced in recent years a very significant interest in research work. Indeed, some systems have been developed to improve this situation. In this paper, we describe the different systems and approaches that were developed and tested in our laboratory to automatically recognize the Amazigh writing, showing the characteristics and results of each. This description will allow us to conduct a comprehensive summary of the various approaches and proposed systems that will help us to launch the outlook for future work.

1. Introduction

Dans le domaine de la reconnaissance automatique des caractères, plusieurs recherches scientifiques ont été effectuées sur le caractère latin, arabe, et autre. Ceci a permis le développement de plusieurs approches de reconnaissance automatique de ces caractères et, par conséquent, des lecteurs optiques pour scanner et reconnaître automatiquement les documents correspondants. Par contre, le caractère amazighe, appelé tifinaghe, est très peu traité. Des approches ont été proposées pour la reconnaissance de ce caractère. Ces approches sont regroupées généralement en grandes classes telles que les approches statistiques (Oulamara, 1988), (Djematen et al., 1997), les approches basées sur les réseaux de neurones (Ait Ouguengay, 2008), (El Yachi et al., 2009), (Bouikhalene et al., 2009) (Es Saady et al., 2011), l'approche syntaxique (Es-Saady et al., 2010), les Modèles de Markov cachés (Amrouch et al., 2010), (Amrouch et al., 2012) et l'approche basée sur la programmation dynamique (El Yachi et al., 2010). Dans cet article, on s'intéresse à présenter un état des lieux et une synthèse présentative et comparative des travaux de recherche scientifique effectués et publiés dans le domaine de la reconnaissance automatique de caractères amazighes imprimés et manuscrits, soit au niveau de notre laboratoire soit ailleurs.

Nous présentons en première partie les principales bases de données de caractères amazighes développées pour pouvoir tester et valider des approches. La seconde

partie est consacrée à la description des différents travaux effectués dans le domaine de la reconnaissance automatique de l'écriture amazighe. Nous présentons dans la troisième partie nos contributions dans ce domaine. Enfin, nous donnons une synthèse détaillée et comparative de ces systèmes tout en précisant les avantages et les limites de chaque système et en lançant un certain nombre de perspectives à développer dans les travaux futurs..

2. Principales bases de caractères amazighs développées

Pour pouvoir expérimenter et valider les différentes approches et systèmes développés, il est primordial de créer des bases de caractères amazighes. Des bases de données d'images de caractères amazighes annotées sont inexistantes. Ce domaine souffre ainsi l'absence d'une base de données de référence qui permet des comparaisons objectives entre les différents systèmes de reconnaissance. Tous les travaux publiés dans ce domaine ont été expérimentés sur des bases de données locales, qui contiennent un nombre restreint de l'alphabet amazighe.

Dans les deux sous-sections suivantes, nous décrivons les deux bases de caractères existantes.

2.1. Base des patterns de la graphie amazighe

C'est une base de patterns de différentes fontes amazighes et de tailles variées proposée par Ait Ouguengay (2009). Elle contient au total 12 polices de caractères et les tailles de 10 points à 28 points pour chaque modèle. Les patterns sont fournis sous forme d'images bitonales de tailles variables. La taille maximale est de 102×129 pixels, tandis que la taille minimale est de 19×2 pixels. Une telle disparité s'explique par le fait que le caractère ya (a) est un petit cercle, et est donc beaucoup plus petit que les autres caractères. Outre le cas particulier du caractère ya (a), la base est constituée des patterns de différentes fontes amazighes et de tailles variées qui ne sont pas normalisées. Le Tableau 1 donne une liste de quelques patterns dans cette base.

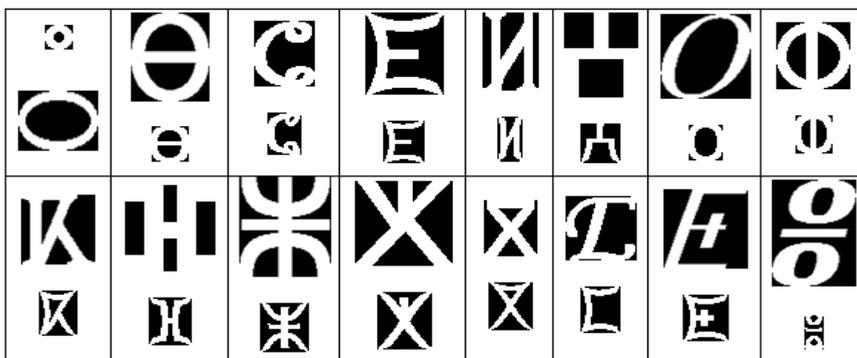


Tableau 1 : Exemples de quelques caractères dans la base des patterns de la graphie amazighe

Dans cette base, la manière dont sont créées les images des patterns ne permet pas la possibilité de renormaliser leur taille en une taille moyenne fixe. En effet, ceci peut être gênant en particulier à cause de la ressemblance des caractères ya (a) et yar (r) qui ne se différencient que par la taille : le caractère ya (a) est un petit cercle, tandis que le caractère yar (r) est un grand cercle. Dans certains cas, on aura une confusion réelle entre des images de ces deux classes.

2.2. Base de caractères manuscrits

C'est une base de données de caractères amazighs manuscrits (A Database for Amazigh Handwritten Character Recognition Research : AMHCD) que nous avons créée et développée au sien de notre Laboratoire IRF-SIC de l'Université Ibn Zohr d'Agadir. La base contient 25740 images de caractères amazighs manuscrits isolés et étiquetés, produites par 60 scripteurs de sexe, d'âge et de fonction différents. Le lecteur peut trouver une description complète et détaillée sur cette base dans (Es Saady et *al.*, 2011).

Jusqu'à présent, la base AMHCD est peu utilisée et explorée pour l'évaluation des systèmes de reconnaissance de l'écriture amazigheD. (2010(Amrouch et *al.*, 2010) (Amrouch et *al.*, 2012a) (Amrouch et *al.*, 2012b). En revanche, elle s'est imposée comme la seule et la première base dans sa catégorie (graphie manuscrite amazighe), grâce à sa taille importante et à sa disponibilité pour les recherches académiques.

Le Tableau 2 présente des exemples des caractères amazighs manuscrits. Chaque caractère est donné sous forme de deux variantes qui correspondent aux deux scripteurs différents.

Caractères amazighs imprimés	Scripteur 1	Scripteur 2	Caractères amazighs Imprimés	Scripteur 1	Scripteur 2

Tableau 2 : Exemples de caractères amazighs manuscrits issus de la base AMHCD (Essaady et al., 2011)

3. Approches de reconnaissance automatique existantes

En comparant au latin, à l'arabe ou au chinois, les recherches sur la reconnaissance automatique d'écriture amazighe n'ont pas atteint la perfection. Autant que nous le sachions, peu de tentatives ont été menées sur la reconnaissance d'écriture amazighe. Dans cette section, nous citons des travaux publiés qui traitent la reconnaissance de l'écriture amazighe.

Parmi les anciennes études qui portent sur la reconnaissance de caractères tifinaghs, on cite, en premier lieu, les travaux d'Oulamara J. (1988). La méthode proposée dans cette référence est une méthode statistique basée sur l'extraction de segments de droite par la transformée de Hough. L'analyse du caractère dans l'espace paramétrique, obtenu par la transformation de Hough, permet d'extraire les caractéristiques spécifiques en association avec un modèle de référence générateur de l'ensemble des caractères de l'alphabet. Un codage original est déduit puis utilisé comme base de construction de la matrice de lecture représentant une forme codée de l'alphabet. L'auteur a obtenu des résultats qui semblent intéressants J. (1988) sur les caractères amazighs imprimés d'une base de caractère locale.

(Djematen et al, 1998) considèrent que la méthode publiée par (Oulamara et al., 1988) n'est pas une technique appropriée pour les caractères amazighs manuscrits puisqu'elle produit des segmentations incorrectes. Pour surmonter la difficulté des caractères présentant des traits inclinés, ces auteurs (Djematen et al, 1997, 1998) proposent une méthode statistique de reconnaissance de caractères berbères manuscrits basée sur la position des points caractéristiques dans le rectangle-enveloppe de l'image du caractère. Après des prétraitements (normalisation bidirectionnelle, le lissage, l'extraction des composantes connexes) sur le caractère,

des primitives sont extraites sur chaque squelette, comme les extrémités, les points, les sommets (points de changements de direction) et les nœuds à 3 et 4 branches. Enfin, la représentation du caractère fournit une description sous forme de lettres utilisant un codage prédéfini. Cette description code les positions des points caractéristiques du caractère dans le rectangle-enveloppe. La reconnaissance consiste à mesurer le degré de ressemblance entre le code élaboré et les codes de référence en utilisant la distance métrique. Les résultats obtenus sont plus ou moins encourageants sur une base de caractères localement définie malgré quelques erreurs qui viennent du module de prétraitement.

Ait Ouguengay (2009) a proposé un réseau de neurones artificiels (RNA) pour la reconnaissance de caractères amazighs. Le réseau de neurones utilisé est un perceptron multicouche à une seule couche cachée. Ce dernier a été entraîné sur une base de données qui contient des patterns de la graphie amazighe de différentes fontes et tailles, créé localement. La simulation du réseau de neurones a été réalisée par le logiciel libre JavaNNS (java neural networks simulator). Les caractéristiques géométriques utilisées sont : les projections horizontales et verticales, les centres de gravité en x et en y, le périmètre, l'aire, la compacité et les moments centraux d'ordre 2. D'après l'auteur, cette approche a donné de bons résultats sur l'ensemble des patterns d'entraînement. Cependant, les résultats de test sont encore loin d'être satisfaisants à cause de la base de test qui est très faible par rapport aux poids de RNA à déterminer.

Pour sa part, El Ayachi et *al.* (2010) propose un système de reconnaissance de l'écriture tiffinagh basé sur les moments invariants et la transformée de Walsh utilisant la programmation dynamique. Le système proposé contient trois parties principales : les prétraitements, l'extraction de caractéristiques et la reconnaissance. Dans le processus de prétraitement, l'image du document numérisé est nettoyée puis elle est segmentée en caractères isolés à l'aide des techniques de l'histogramme. Dans le processus d'extraction de caractéristiques, les moments invariants et les coefficients de Walsh sont calculés sur les caractères segmentés. La programmation dynamique est adoptée dans l'étape de reconnaissance. Les tests ont été faits sur plusieurs images d'écriture amazighe. D'après les auteurs, les résultats expérimentaux montrent que la méthode de la reconnaissance utilisant des moments invariants donne de meilleurs résultats par rapport à la méthode fondée sur la transformée de Walsh en termes de taux de reconnaissance, de taux d'erreur et de temps de calcul. Plus récemment, les mêmes auteurs (2011) ont proposé dans un réseau de neurones multicouches avec les mêmes caractéristiques utilisées précédemment. Les résultats trouvés avec un réseau de neurones d'une seule couche cachée sont meilleurs que ceux obtenus avec la programmation dynamique. De plus, le taux de reconnaissance obtenu en utilisant une seule couche cachée est plus élevé que celui obtenu avec deux ou trois couches cachées.

Dans la section suivante, nous présentons nos contributions en terme de conception et de développement de nouvelles approches de reconnaissance automatique de caractères amazighs imprimés ou manuscrits en les testant sur la base de caractères volumineuse et riche (A Database for Amazigh Handwritten Character Recognition Research :AMHCD) (Es Saady et *al.*, 2011).

4. Nos systèmes de reconnaissance automatique proposés

4.1. Approches Markoviennes

Dans (Amrouch et *al.*, 2010, 2012a), nous avons proposé une approche basée sur les modèles de Markov cachés de type modèle discriminant DM-HMM qui s'intéresse aux problèmes de caractères isolés. Ce type de modélisation est largement utilisé dans le domaine de la reconnaissance de la parole. Cette modélisation est aussi efficace pour reconnaître une forme entachée d'incertitude et d'aspect de dynamisme comme le caractère amazigh. Un processus Markovien a mis en oeuvre des modèles probabilistes spécifiques dans le but de gérer l'incertitude et le manque d'informations qui entachent les formes à reconnaître. Après des prétraitements sur l'image du caractère amazigh, le système fait recourt à des primitives directionnelles dans la génération des séquences d'observations. Ces observations sont obtenues à l'aide de la technique des fenêtres glissantes opérant sur la transformée standard de Hough des images de caractères. Les séquences d'observations obtenues sont utilisées pour entraîner les modèles HMMs initiaux des caractères lors de la phase d'apprentissage ; chaque modèle utilise les échantillons de sa classe. Par la suite, nous avons utilisé le classifieur Forward pour reconnaître le caractère. En effet, cette approche consiste à associer un ou plusieurs modèles par classe. De ce fait, la reconnaissance s'effectue en estimant les probabilités d'émission de la suite d'observations O de la forme à reconnaître par les différents modèles préalablement construits. La forme à reconnaître est affectée à la classe dont le modèle qui maximise la probabilité. Cette approche est pratiquement utilisée dans le cas où le nombre de classes à reconnaître est relativement limité, c'est-à-dire au vocabulaire limité comme l'alphabet amazigh. Toutefois, elle devient coûteuse en temps de calcul et espace mémoire quand ce nombre dépasse le millier, puisque chaque classe possède au moins un modèle qui lui est propre.

Nous avons évalué les performances de notre système sur la base AMHCD des caractères amazighes manuscrits isolées (Es Saady et *al.*, 2011), avec deux variantes. La première adopte la modélisation discrète des probabilités d'émission, quant à la seconde, elle utilise les HMMs continus. Le tableau 3 ci-dessous présente les différents taux de reconnaissance en utilisant le modèle discret et continu selon le nombre d'états de modèle et de la taille de la base de caractère.

Topologie de modèle utilisée	HMMs Discrets		HMMs Continus	
	Taille de la base	Taux de Recon	Taille de la base	Taux de Recon
Modèle à 14 états	2220 caractères	90.04 %	-	-
Modèle mono et 2 gaussiens à 6 états	-	-	25740 caractères	96,21%
Modèle mono et 2 gaussiens à 10 états	-	-	25740 caractères	96 , 88%
Modèle mono et 2 gaussiens à 14 états	-	-	25740 caractères	97 , 89%

Tableau 3 : Résultats de reconnaissance du système dans le cas discret et continu

Nous estimons que les erreurs de la variante discrète proviennent essentiellement de: (1) la modélisation discrète utilisée, à ce niveau, on a recours au risque de la perte des informations ; (2) à la taille des données utilisées pour créer la base des modèles de référence.

Pour valider cette hypothèse et remédier à cette défaillance, nous avons augmenté la taille des données utilisées et remplacé les modélisations discrètes par les HMMs continus.

L'augmentation de la taille des données d'apprentissage et la modélisation des densités des probabilités par les gaussiennes ont contribué à diminuer le taux d'erreur commis par notre système. Nous avons passé d'un taux d'erreur global de 9,6% (première expérience avec HMM discrets) à un taux de 2,11% c'est-à-dire un gain d'un facteur de 7,49%. De ce fait, les résultats obtenus dans le cas continu sont meilleurs que ceux obtenus dans le cas discret.

Dans (Amrouch et *al.*, 2012b, 2012c), nous avons proposé un autre système pour la reconnaissance de caractères Tifinaghs imprimés, basé sur une nouvelle approche qui exploite les caractéristiques et les spécificités morphologiques de la langue amazighe. La solution apportée adopte une modélisation markovienne de type chemin discriminant (DP-HMM), optimisée par des algorithmes fondés sur la programmation dynamique S.K. (1996 ; Casey, E. (1996). L'approche s'appuie sur la proposition d'une nouvelle liste des segments, qui se compose d'un ensemble de traits fondamentaux constituant les caractères amazighs. Ceci permet de mieux exploiter la redondance de ces traits dans les tracés des lettres amazighes. La description de la structure des caractères repose sur ces éléments. En effet, les caractéristiques exploitées sont extraites des tracés des caractères par une technique de localisation implicite des segments qui le composent. Pour ce faire, nous avons utilisé les points d'intérêts des squelettes. Dans la phase de l'apprentissage, un seul modèle HMM global construit et entraîné sur les éléments du vocabulaire proposé par des primitives structurelles et géométriques. Chaque chemin au long de ce treillis représente une séquence de segments, qui constitue un caractère de

l'alphabet tifinagh. La reconnaissance s'effectue en décodant dynamiquement le chemin optimal suivant le critère de maximum de vraisemblance.

Pour valider le système proposé, nous avons effectué des expérimentations significatives sur la base de données de patterns de la graphie amazighe (Essaady et al., 2011). Plusieurs tests ont été effectués pour évaluer le taux de reconnaissance du système en fonction du : nombre d'états et du nombre de mélanges de gaussiennes. Le Tableau 4 ci-dessous présente les résultats obtenus de ces tests sur cette base.

Nombre d'états	3	5
Nombre de gaussienne	1-2-3	1-2-3
Taux de reconnaissance	98 , 41%	98 , 76%

Tableau 4 : Taux de reconnaissance sur BDI

Ces résultats montrent un taux d'erreur de 1,24% avec un modèle de topologie de 5 états. Nous estimons que les erreurs de reconnaissance sont attribuées, d'une part, aux méthodes utilisées pour la pré-classification et à la détection des points d'intérêts et, d'autre part, à l'insuffisance des caractéristiques utilisées pour mieux décrire chaque segment. En outre, cette faible erreur provient aussi de la déformation de certains caractères dans certaines fontes, notamment dans les fonts « Tassafut » et « Taromeit ». Nous constatons que le nombre de gaussiennes utilisées n'influence pas les résultats alors que son augmentation implique un nombre important de paramètres à calculer. Cependant, le choix de la topologie influence directement les résultats. Par conséquent, l'augmentation de nombre d'états augmente le taux de reconnaissance de système.

4.2. Approches syntaxiques

Dans (Essaady et al., 2010), nous avons présenté un système automatique de reconnaissance de caractères amazighs imprimés, basé sur une approche syntaxique utilisant les automates finis. Après des prétraitements sur l'image du caractère, des algorithmes appropriés sur le squelette de caractère permettent de construire la chaîne représentative du caractère à partir du codage de Freeman. La chaîne reconstruite est utilisée à l'entrée d'un automate fini qui reconnaît les caractères amazighs. Cet automate global a été construit à partir des automates de reconnaissance spécifique à chaque caractère amazigh. Nous avons testé notre application sur une base de caractères amazighs imprimés que nous avons créée. Nous avons obtenu des résultats encourageants. En effet, sur les 630 caractères lus, 589 ont été reconnus, soit un taux de reconnaissance de 93,49%. Le Tableau 5 ci-dessous présente les taux des mauvaises affectations et mauvais rejets. Ces erreurs proviennent de la forme de certains caractères non reconnus, dont le squelette comporte plus de segments non orthogonaux. En effet, la méthode de reconnaissance est basée sur une vectorisation du squelette du caractère à reconnaître. Donc, une erreur de vectorisation va forcément entraîner une erreur dans la description du caractère. En fait, le principal inconvénient de cette méthode est la sensibilité du squelette au bruit.

Taux d'affectations à tort	2,28 %
Taux de rejets à tort	4,23 %

Tableau 5 : Pourcentages des erreurs de reconnaissance

4.3. Approches neuronales

El Ayachi et *al.* (2010) proposent un système de reconnaissance de l'écriture tiffinagh basé sur les réseaux de neurones multi couches avec les mêmes caractéristiques utilisées précédemment. Les résultats trouvés avec un réseau de neurones d'une seule couche cachée sont meilleurs que ceux obtenus avec la programmation dynamique. De plus, le taux de reconnaissance obtenu en utilisant une seule couche cachée est plus élevé que celui obtenu avec deux ou trois couches cachées.

Dans (Essaady et *al.*, 2011b), et pour améliorer les résultats du précédent système et d'avoir un système complet qui reconnaît la totalité des caractères amazighs imprimés et manuscrits, nous avons présenté un système de reconnaissance automatique du texte amazigh, basé sur les réseaux de neurones multicouches. L'approche proposée a recours à des primitives statistiques en se basant sur la position des lignes centrales du caractère et les techniques de fenêtres glissantes.

Dans ce système et dans un premier temps, nous avons utilisé la ligne centrale horizontale du caractère pour extraire un ensemble de caractéristiques de densité basées sur cette ligne. Dans un second temps, nous avons proposé une amélioration de ce système de reconnaissance en ajoutant d'autres caractéristiques de densité basées sur la ligne centrale verticale du caractère afin d'exploiter la similarité de plusieurs caractères amazighs par rapport à la ligne centrale verticale du caractère. Les différentes variantes ont été testées et évaluées sur deux bases : la base des patterns de la graphie amazighe (Ait ougangay, 2009) et la base AMHCD des caractères amazighs manuscrits (Essaady et *al.*, 2011a). En effet, nous avons montré les résultats de reconnaissance obtenus en fonction de l'intégration des caractéristiques dépendantes et indépendantes à la ligne centrale horizontale du caractère. Enfin, nous avons présenté les résultats obtenus par la version améliorée et leur comparaison avec ceux obtenus par la première variante du système.

Nous avons aussi utilisé des techniques de validation croisée pour l'évaluation des résultats de reconnaissance. La validation croisée est une méthode d'estimation de la fiabilité des résultats, fondée sur une technique d'échantillonnage R. (1995).

Le Tableau 6, ci-dessous, présente les résultats du système proposé en utilisant la validation croisée 10 fois sur la base des patterns de la graphie amazighe et sur la base de caractères manuscrits.

Les caractéristiques intégrées	Base des patterns de la graphie amazighe		Base de caractères amazighs manuscrits	
	Taille de la base	Taux de Recon	Taille de la base	Taux de Recon
Caractéristiques indépendantes de la ligne centrale	19437 caractères	88.68 %	20150 caractères	84.49 %
Caractéristiques dépendantes et indépendantes de la ligne centrale horizontale	19437 caractères	98.49 %	20150 caractères	92.23 %
Caractéristiques dépendantes et indépendantes de la ligne centrale horizontale et verticale	19437 caractères	99.28 %	20150 caractères	96.32 %

Tableau 6 : Résultats de reconnaissance du système amélioré en fonction des caractéristiques intégrées en utilisant la validation croisée 10 fois

Pour la base des patterns de la graphie amazighe, le taux de reconnaissance est 88,68% lors de l'utilisation seulement des caractéristiques indépendantes de la ligne centrale horizontale et augmente à 98,49% lors de l'ajout des caractéristiques basées sur la position de la ligne centrale horizontale. Ce dernier taux s'élève aussi à 99,28% lors de l'ajout des caractéristiques basées sur la position de la ligne centrale verticale.

Pour la base de caractères amazighs manuscrits, le taux augmente de 84,49% à 92,23% lors de l'ajout des caractéristiques basées sur la position de la ligne centrale horizontale et monte à 96,32% lors de l'ajout des caractéristiques basées sur la position de la ligne centrale verticale.

Les erreurs sont principalement dues à une grande similarité morphologique entre certains caractères amazighs et, parfois, sur des fontes différentes.

En comparant les différents taux de reconnaissance obtenus par le système en fonction de caractéristiques intégrées, nous constatons une amélioration due à l'intégration des caractéristiques basées sur la position des deux lignes centrales (verticale et horizontale). Cela confirme que ces caractéristiques offrent une amélioration significative à la performance de reconnaissance.

5. Synthèse comparative

Le Tableau 7 récapitule les principales caractéristiques, les techniques et les résultats obtenus par des approches et les systèmes de reconnaissance de l'écriture amazighe que nous avons proposés.

Auteurs	Primitives	Modélisation	Résultats
(Amrouch et al., 2010, 2012a)	- primitives directionnelles (Transformée de Hough)	- HMM continu et discret 14 -états - topologies de type gauche droit. - apprentissage le critère (MLE) - Classification FORWARD	- 90.4 % en cas discret. - 97, 89% en cas continu. - Base (AMHCD). - Base d'apprentissage (2/3). - Base de test (1/3).
(Amrouch et al., 2012b, 2012c)	- primitives structurales (diamètre, excentricité, étendue, Centre de masse, orientation, longueur d'axe minimal et principal, moment d'ordre 1 et 2)	- HMM continu. - topologies de type linéaire et ergodique - Apprentissage Baum-welsh - Classification Viterbi	- 98,76% la base de données de patterns de la graphie amazighe. Base d'apprentissage (2/3) Base de test (1/3)
(Essaady et al., 2010)	- Codage de Freeman (suivi de squelette)	- réseaux des automates. - (Grammaire formelle)	- 93,49% - selon les données utilisées.
(Essaady et al., 2011c)	- primitives statistiques et caractéristiques dépendantes et indépendantes de la ligne centrale horizontale et verticale - fenêtres glissantes horizontales et verticales	- Réseaux de neurones multicouches. - couche d'entrée 95 neurones - couche de sortie 31 neurones - couche cache (nb d'entrées + nb de sorties)/2	- 99.28 % Base des patterns de la graphie amazighe (19437 caractères) - 96.32 % Base AMHCD (20150 caractères)

		- Apprentissage de Heb ($\eta=0,2$, taux=0,3, itéra=1000)	
(Elaychi et al., 2010)	- les moments invariants et la transformée de Walsh utilisant la programmation dynamique.	- La programmation dynamique	- Taux de reconnaissance très intéressant
(Elaychi et al., 2011)	- un réseau de neurones multi couches avec les mêmes caractéristiques utilisées précédemment.	- Programmation dynamique	- Résultats très intéressants
(Elaychi et al., 2011b)	- Les caractéristiques géométriques : les projections horizontales et verticales, les centres de gravité en x et en y, le périmètre, l'aire, la compacité et les moments centraux d'ordre 2.	- réseau de neurones artificiels (RNA)	- Résultats très intéressants
(Djematen et al., 1998)	- les extrémités, les points, les sommets (points de changements de direction) et les nœuds à 3 et 4 branches.	- méthode statistique : mesurer les degrés de vraisemblance	- Résultats plus ou moins intéressants

Tableau 7 : synthèse des approches et systèmes proposés

Après avoir étudié et comparé les différentes approches, nous constatons que la meilleure et la performante approche, quant au temps de calcul, espace mémoire et taux de reconnaissance, est l'approche basée sur les primitives statistiques et les fenêtres glissantes dans la phase de prétraitement et les réseaux de neurones dans la phase de classification (Essaady et al., 2011c).

6. Conclusion et perspectives

Nous avons présenté dans ce papier des travaux traitant la reconnaissance automatique des caractères amazighs. L'objectif est de faire une synthèse des travaux effectués sur ce sujet. Ces approches présentent un certain nombre de limites qui proviennent à la fois du module de prétraitement et des caractéristiques prises dans la phase d'apprentissage. En plus, les bases de caractères utilisées dans les tests restent parfois faibles et parfois non standards. Par conséquent, des travaux de recherche futurs doivent apporter des améliorations, d'un côté, sur ces approches et, d'un autre côté, développer d'autres systèmes qui répondent aux attentes. Et parmi nos travaux futurs, nous allons proposer des systèmes hybrides qui utilisent des primitives de nature différente en combinant ces approches dans le processus de traitement. Ce qui permettra de profiter *a priori* des avantages de chacune des approches tout en évitant les principaux inconvénients.

7. Références bibliographiques

- Ait Ouguengay Y., Taalabi M. (2009), Elaboration d'un réseau de neurones artificiels pour la reconnaissance optique de la graphie amazighe: Phase d'apprentissage, *Systèmes intelligents-Théories et applications, Paris : Europa, cop. (impr. au Maroc)*, ISBN-102909285553, Avril 2009.
- Amrouch M., Rachidi A., El Yassa M., Mammass D. (2010), "Handwritten Amazigh Character Recognition Based On Hidden Markov Models", *ICGST-GVIP Journal*, vol.10, Issue 5, pp.11-18, December 2010.
- Amrouch M., Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2012a), "Handwritten Amazigh Character Recognition System Based on Continuous HMMs and Directional Features", *IJMER journal*, Vol.2, Issue 2, pp.436-441, Mar.-Apr. 2012.
- Amrouch M., Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2012b), "A New Approach Based on Strokes for Printed Tifinagh Character Recognition Using Discriminating Path-HMM", *IRECOS journal*, Vol.7, N°.2, Mars 2012.
- Amrouch M., Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2012c), "A Novel Feature Set for Recognition of Printed Amazigh Text Using Maximum Deviation and HMM", *IJCA journal*, Vol.44, N°.12, pp.23-30, April 2012.
- Casey R.G. and Lecolinet E. (1996), A survey of methods and strategies in character segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol.18,N°.7, pp.690-706, Jul 1996.
- Djematen A., Taconet B. and Zahour A. (1997), "A Geometrical Method for Printing and Handwritten Berber Character Recognition", *ICDAR'97*, pp.564, 1997.
- Djematen A., Taconet B. and Zahour A. (1998), « Une méthode statistique pour la reconnaissance de caractères berbères manuscrits », *CIFED'98*, pp.170-178, 1998.

- El Ayachi R., Moro K., Fakir M. and Bouikhalene B. (2010), "On the Recognition of Tifinaghe Scripts", *Journal of Theoretical and Applied Information Technology*, vol.20 (2), pp.61-66, 2010.
- El Ayachi R., Moro K., Fakir M. and Bouikhalene B. (2011), "Recognition of Tifinaghe Characters Using a Multilayer Neural Network", *International Journal Of Image Processing (IJIP)*, vol. 5, Issue 2, 2011.
- Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2011), AMHCD: A Database for Amazigh Handwritten Character Recognition Research. *International Journal of Computer Applications* , Vol.27, N°.4, pp:44-48, August 2011. Published by Foundation of Computer Science, New York, USA.
- Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2010), "Printed Amazigh Character Recognition by a Syntactic Approach using Finite Automata", *ICGST-GVIP Journal*, vol.10, Issue 2, pp.1-8, 2010.
- Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2011a), "AMHCD: A Database for Amazigh Handwritten Character Recognition Research", *International Journal of Computer Applications*, vol.27 (4), pp.44-48, published by Foundation of Computer Science, New York, August 2011.
- Es Saady Y., Rachidi A., El Yassa M., Mammass D. (2011b), "Amazigh Handwritten Character Recognition based on Horizontal and Vertical Centerline of Character", *International Journal of Advanced Science and Technology*, vol.33, pp.33-50, August, 2011.
- Kim H.J., Jung J.W. and Kim S.K. (1996), On-line Chinese character recognition using ARTbased stroke classification. *Pattern Recognition Letters*, Vol.17, N°.12, pp.1311–1322, 1996.
- Kohavi R. (1995), "A study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection", *International Joint Conference on Artificial Intelligence (IJCAI)*, 1995.
- Ly Van B. (2005), "Réalisation d'un Système de Vérification de Signature Manuscrite En-ligne Indépendant de la Plateforme d'Acquisition", *Thèse de doctorat de l'Institut National des Télécommunications*, Décembre 2005.
- Oulamara A. and Duvernoy J. (1988), "An application of the Hough transform to automatic recognition of Berber characters", *Signal Processing*, vol.14, pp.79-90, 1988.