

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE



UNIVERSITE MOHAMED BOUDIAF - M'SILA
FACULTE DES MATHÉMATIQUES ET
DE L'INFORMATIQUE



DEPARTEMENT D'INFORMATIQUE

MEMOIRE de fin d'étude

Présenté pour l'obtention du diplôme de MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : Systèmes d'Informations Avancés

Par: Baza Halima Saadia

SUJET

La reconnaissance d'écriture arabe manuscrite

Soutenu publiquement le : / /2016 devant le jury composé de :

.....	Université de M'sila	Président
Kamel Mohamed	Université de M'sila	Rapporteur
.....	Université de M'sila	Examineur
.....	Université de M'sila	Examineur

Promotion : 2015 /2016

TABLE DES MATIERES

Introduction générale.....	6
<u>Chapitre I : Reconnaissance de formes</u>	
1. Introduction:	8
2. Reconnaissance des formes :.....	8
2.1. La reconnaissance d'objet:	8
2.2. Similarité des formes et les taches de RF :.....	8
2.2.1. Tache de classification :.....	9
2.2.2. Taches de régression :	9
2.2.3. Taches de description :.....	10
2.3. Classes, formes et caracteristiques :.....	10
2.3.1. Classes :.....	10
2.3.2. Formes :	10
2.3.3. Caractéristiques:.....	11
2.4. Les approches du RF:	11
2.5. Projet du RF :.....	13
2.5.1. Taches du projet :	13
2.5.2. Apprentissage et test :.....	14
2.5.3. Les logiciels de la reconnaissance:.....	14
3. La reconnaissance d'écriture:.....	15
3.1. Les phases d'un système de reconnaissance d'écriture :.....	15
3.1.1. Acquisition d'image :.....	15
3.1.2. Prétraitements et normalisations :.....	16
3.1.3. Segmentation :	21
3.1.4. Extraction des caractéristiques:.....	23
3.1.5. Post-traitements :.....	24
4. La reconnaissance de l'écriture arabe manuscrite :	24
4.1. Problématique :	24
4.1.1. Présentation de la langue arabe :	24

4.1.2. Difficultés inhérentes à la reconnaissance de l'écriture arabe	25
5. Conclusion :	29
<u>Chapitre II: Les systèmes de classification.</u>	
1. INTRODUCTION :.....	31
2. Définition formelle d'un classificateur :.....	31
2.1. Classificateur de type classe :.....	32
2.2. Classificateur de type rang :.....	32
2.3. Classificateur de type mesure :	32
3. Les réseaux de neurones :.....	33
3.1. Définition :.....	33
3.2. Neurone simple :.....	33
3.2.1. Le neurone de Mac Culloch-Pitts :.....	33
3.2.2. Le Perceptron :.....	34
3.3. Interprétation mathématique :.....	34
3.4. Types et architectures des RNA :.....	35
3.4.1. Le Perceptron Multicouches :.....	35
3.4.2. Fonctions à bases radiales :	37
3.5. L'apprentissage :.....	38
3.5.1. La loi d'apprentissage de Hebb [HEB49] :	38
3.5.2. La règle d'apprentissage du Perceptron :	39
3.5.3. Retro-propagation du gradient :	39
3.6. Avantages et inconvénients des réseaux neuronaux :	40
3.6.1. Avantages de l'approche neuronale :.....	40
3.6.2. Inconvénients de l'approche neuronale :	40
4. K plus proches voisins :.....	41
4.1. Formalisation mathématique :.....	41
4.2. ALGORITHME :.....	42
5. Modèles de markov caches :	43
5.1. Définitions :.....	43
5.1.1. Chaîne de Markov :	43
5.1.2. Chaîne stationnaire :	43
5.2. Un modèle de markov cache :.....	44

5.3. <i>Interprétation mathématique</i> :.....	44
5.4. <i>Les types des HMMs</i> :.....	45
5.4.1. <i>HMM ergodique</i> :.....	45
5.4.2. <i>HMM gauche-droite</i> :.....	45
6. <i>Conclusion</i> :.....	46
<u>Chapitre III : Conception et résultats.</u>	

1. <i>Introduction</i> :.....	48
2. <i>Conception de la solution proposée</i> :.....	48
2.1. <i>Prétraitement</i> :.....	48
2.1.1. <i>Binarisation</i> :.....	49
2.1.2. <i>Squelettisation</i> :.....	49
2.1.3. <i>Centrage</i> :.....	50
2.1.4. <i>Extraction des caractéristiques</i> :.....	50
2.2. <i>Apprentissage et classification</i> :.....	51
3. <i>Base des caractères et outils de développement</i> :.....	51
3.1. <i>Base des Caractères</i> :.....	51
3.2. <i>Mathworks MATLAB</i> :.....	52
4. <i>Présentation de l'application</i> :.....	53
4.1. <i>Les étapes de reconnaissance d'un caractère</i> :.....	53
4.1.1. <i>Chargement de l'image à partir de la base des caractères</i> :.....	53
4.1.2. <i>La binarisation</i> :.....	54
4.1.3. <i>Squelettisation</i> :.....	54
4.1.4. <i>Centrage</i> :.....	54
4.1.5. <i>Extraction des caractéristiques</i> :.....	55
4.1.6. <i>Reconnaissance</i> :.....	55
5. <i>Test et validation</i> :.....	55
5.1. <i>Classification</i> :.....	55
5.1.1. <i>Le premier RNA</i> :.....	56
5.1.2. <i>Le deuxième RNA</i> :.....	56
5.1.3. <i>Le troisième RNA</i> :.....	57
5.1.4. <i>Le quatrième RNA</i> :.....	58
5.1.5. <i>Le cinquième RNA</i> :.....	59

6. Combinaison :	60
7. Conclusion :	61
Conclusion et perspectives	61
Bibliographie	63
Liste des figures	64
Liste des tableaux	65

Introduction général

Les techniques liées au traitement de l'information connaissent actuellement un développement très actif en liaison avec l'informatique et présentent un potentiel de plus en plus important dans le domaine de l'interface homme machine. L'écriture restera l'un des grands fondements des civilisations et le mode par excellence de conservation et de transmission du savoir. La reconnaissance de l'écriture est un domaine vaste qui constitue un sous ensemble des systèmes de la reconnaissance des formes (RF). Ces systèmes sont la première étape d'un processus de compréhension de notre univers dans le cadre global de la communication homme machine. La reconnaissance de l'écriture manuscrite ou imprimée reste encore un sujet de recherche et d'expérimentation, le problème n'est pas encore entièrement résolu bien que l'on sache atteindre des taux assez élevés dans certaines applications pour lesquelles soit le vocabulaire est limité, soit la fonte est unique ou en nombre restreint.

Il existe cependant plusieurs domaines pour lesquels la reconnaissance de l'écriture est appliquée avec un certain succès : le tri automatique du courrier, le traitement automatique de dossiers administratifs, des formulaires d'enquêtes, ou encore l'enregistrement des chèques bancaires. La reconnaissance de l'écriture manuscrite est beaucoup plus complexe que celle de l'écriture imprimée due à son extrême variabilité: variabilité des formes, des espacements entre mots et caractères, fluctuation des lignes.

Les travaux de recherche en reconnaissance de l'écriture arabe manuscrite, bien que moins avancés que pour d'autres langues, deviennent plus intensifs qu'avant.

Après cette introduction, dans le premier chapitre, nous présenterons rapidement les notions de bases de la reconnaissance des formes, la reconnaissance de l'écriture et la reconnaissance de l'écriture arabe manuscrite.

Dans le deuxième chapitre, nous présenterons un ensemble de classificateurs couramment utilisés en reconnaissance de l'écriture arabe.

Dans le troisième chapitre nous présenterons notre travail sur la reconnaissance des caractères arabes manuscrits, L'originalité de notre travail consiste à construire un classificateur neuronale composé de cinq réseaux de neurones, avant d'utiliser une stratégie pour les combiner.

**CONCLUSION
ET
PERSPECTIVES**

Le domaine de la reconnaissance de la langue arabe est très vaste et riche en connaissance, que ce soit sur la langue elle-même ou ses différentes tournures et ses ambiguïtés.

Le problème de la reconnaissance des caractères arabes est un axe de recherche en plein d'expansion et attire les chercheurs.

Ce mémoire porte sur un système intelligent de classification basée sur les Réseaux de Neurons pour la reconnaissance des caractères arabes manuscrits avec des scripteurs multiples.

Un système de classification a été réalisé pour un ensemble de caractères réduit et fixé, et pour des réseaux neuronaux multiples. Ce système obtient des performances meilleures qu'avec un seul réseau comme montré dans le dernier chapitre.

Par ce travail nous espérons que nous avons couvert une partie concernant le domaine de recherche de la combinaison des classificateurs appliquée sur l'écriture arabe manuscrite, et pu contribuer à l'évolution des recherches, malgré que les efforts de nos jours s'intensifient dans ce domaine et chaque jour de nouveaux articles sont publiés, traitant du sujet.

Dans l'avenir il est possible d'intégrer un ou plusieurs classificateurs basé sur K-Plus Proches Voisins dans le système de classification et essayer de bénéficier les puissances des K-Plus Proches Voisins montrées dans les travaux d'aujourd'hui.

Nous espérons dans l'avenir pouvoir intégrer un groupe de recherche dans ce domaine pour pouvoir contribuer dans la recherche ainsi que développer notre algorithme pour couvrir plus de caractères arabes et plus de scripteurs et contribuer dans l'avancement des recherches dans le domaine de reconnaissance de l'écriture arabe manuscrite.

BIBLIOGRAPHIE

[BAD95]	B. Al-Badr and S.A. Mahmoud, "Survey and bibliography of Arabic optical text recognition," <i>Signal processing</i> , vol. 41, pp. 49-77, 1995.
[BEL97]	A. Belaid. (1997) Analyse de documents: de l'image à la représentation par les normes de codage.
[BEN99b]	A. Bennis, A. Zahour, and B. Taconet, "Extraction des lignes d'un texte manuscrit arabe," <i>Vision interface</i> , 1999.
[CAS96]	R.G. Casey and E. Lecolinet, "A survey of methods and strategies in character segmentation," <i>IEEE Transactions on pattern analysis and machine intelligence</i> , vol. 18, no. 7, July 1996.
[CRO02]	James L. Crowley, "Reconnaissance de formes dans les images", DESS GI - Option IHM, 22, janvier 2002.
[DUP04]	Xavier Dupré, "Contributions à la reconnaissance de l'écriture cursive à l'aide de modèles de Markov cachés," Paris V, 2004.
[HAL04]	Khalid Hallouli, "Reconnaissance de caractères par méthodes markoviennes et réseaux baésiens," Paris, Thèse de doctorat 2004.
[HAY94]	S. Haykin, <i>Neural Networks: A Comprehensive Foundation</i> . New York: Macmillan, 1994.
[HEB49]	D. Hebb, <i>The Organization of Behavior</i> . New York: Wiley, 1949.
[MAR01]	J. P. Marques de Sa, <i>Pattern Recognition: Concepts, Methods and Applications</i> .: Springer, 2001.
[MCC43]	W. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," <i>Bulletin of Mathematical Biophysics</i> , vol. 5, pp. 115-133, 1943.
[MEN08]	Farès Menasri, "Contributions à la reconnaissance de l'écriture arabe manuscrite," UNIVERSITE PARIS DESCARTES, Thèse de doctorat 2008.
[NIG93]	A. Nigrin, <i>Networks for Pattern Recognition</i> . Cambridge: MA: The MIT Press, 1993.
[OLI02]	J. J. Oliveira, J. de Carvalho, C. Freitas, and R. Sabourin, <i>Feature sets evaluation for handwritten word recognition</i> ., 2002.
[PEC02]	M. Pechwitz, S. Snoussi Maddouri, V. Maergner, N. Ellouze, and H. Amiri, "IFN/ENIT database of handwritten arabic words," <i>CIFED</i> , 2002.
[ROS58]	F. Rosenblatt, "The Perceptron: a probabilistic model for information storage and organization in the brain," <i>Psychological Review</i> , vol. 65, pp. 386-408, 1958.
[XUL92]	L. Xu, A. Krzyzak, and C. Suen, "Methods of combining multiple classifiers and their applications to handwriting recognition," <i>IEEE Transactions on Systems, Man, and Cybernetics</i> , vol. 22, no. 3, pp. 418-435, 1992.

RESUME

Le présent travail porte sur une étude concernant le domaine de reconnaissance des formes (RF) appliqué sur l'écriture arabe manuscrite, il s'agit d'abord de faire une étude générale sur la RF, puis d'étudier les systèmes de classification couramment utilisés dans ce domaine, puis de faire une étude bibliographique sur les systèmes existants et les différentes recherches effectuées sur la reconnaissance d'écriture arabe. Pour enfin proposer un système intelligent de classification pour la reconnaissance des caractères arabes manuscrits à l'aide de réseaux de neurones.

MOTS-CLES : *RF, Reconnaissance des caractères arabes, caractère manuscrit, extraction des caractéristiques, classificateur, RNA, pré-traitement.*

التلخيص

هذا العمل يندرج ضمن دراسة مجال التعرف على الأشكال (RF) المطبقة على النصوص العربية المكتوبة يدويا ، في البداية سنقوم بدراسة عامة فيما يخص التعرف على الأشكال ، ثم دراسة مجموعة من أنظمة التصنيف المستخدمة عادة في هذا المجال ، ثم القيام بدراسة على الأنظمة القائمة والبحوث المختلفة في مجال التعرف على الكتابة العربية، وأخيرا اقتراح نظام ذكي للتعرف على الحروف العربية المكتوبة يدويا باستخدام شبكات الخلايا العصبية الاصطناعية.

الكلمات المفتاحية: التعرف الضوئي على الحروف، التعرف على الحروف العربية ، حرف مكتوب يدويا، واستخراج الخصائص ، المصنف ، المعالجة القبلية.

Abstract

This work is a study of Pattern Recognition (PR) field applied to the Arabic handwriting, first, we will make a general study of PR systems, then we study the classification systems commonly used in this field, then do a literature review on existing systems and various research on the Arabic writing recognition, finally, we will propose an intelligent classification system for *handwritten Arabic characters* using Artificial neural networks.

KEYWORDS: *PR, Arabic characters Recognition, handwritten character, feature extraction, classifier, ANN, pre-treatment.*