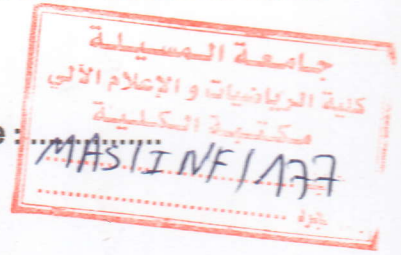




N° d'ordre :



UNIVERSITE DE M'SILA
FACULTE DES MATHÉMATIQUES ET DE L'INFORMATIQUE
Département d'Informatique

MEMOIRE de fin d'étude
Présenté pour l'obtention du diplôme de MASTER
Domaine : Mathématiques et Informatique
Filière : Informatique
Spécialité : Technologie de l'information et de communication
Par: RAHALI Anfel

SUJET

**RECONNAISSANCE AUTOMATIQUE DES
CARACTERES ARABES MANUSCRIT AVEC LES
MOMENTS DE ZERNIKE**

Soutenu publiquement le : 16/06/2015 devant le jury composé de :

M.MOHAMED Kamel
Mdm. ASSAS Warda
M.AMRI

Université de M'sila
Université de M'sila
Université de M'sila

Président
Rapporteur
Examineur

Promotion : 2014 / 2015

INTRODUCTION GENERALE

La reconnaissance de l'écriture est un sujet de recherche populaire dans la reconnaissance des formes. Dont le but est de développer un système qui se rapproche le plus de l'être humain dans sa capacité de lire, et de rendre la communication homme-machine plus facile et plus souple. La reconnaissance de l'écriture a de nombreuses applications, notamment en bureautique et tri automatique du courrier, pour l'enregistrement et la vérification des chèques bancaires, en publication assistée par ordinateur (PAO) pour faciliter la composition à partir d'une sélection de plusieurs documents et le traitement automatique de dossiers administratifs.

L'écriture manuscrite demeure aujourd'hui l'un des moyens de communication les plus simples et les plus expressifs, permettant d'exprimer l'identité et la culture d'un individu. L'écriture manuscrite a ainsi su s'imposer comme un fondement de nombreuses civilisations. Aujourd'hui, malgré l'avènement des nouvelles technologies, elle reste un moyen de communication incontournable. La reconnaissance de l'écriture manuscrite demeure un sujet de recherche vivace et vaste les travaux de recherches sur l'écriture arabe sont moins nombreux en comparaison avec d'autres types d'écriture (le latin, le japonais...). En plus la cursivité de l'écriture arabe montre une complexité de la morphologie des caractères. Ce problème engendre une forte inertie à différents niveaux notamment :

- Le choix de primitives pertinentes décrivant la variabilité de la morphologie des caractères, sachant que certaines caractéristiques topologiques sont sensibles à la dégradation, notamment les points diacritiques et les boucles
- La nécessité d'une modélisation robuste et une méthode d'apprentissage efficace pour prendre en considération toutes les variations morphologiques de l'écriture arabe

Les travaux effectués dans le cadre de ce mémoire portent sur la reconnaissance du texte arabe manuscrit hors-ligne. L'objectif est de présenter une architecture d'OCR (Optical Character Recognition System) par les moments invariants pour extraire les paramètres de discrimination et les réseaux de nouerons et le support vectors machine comme des classifieurs. Ce système se base sur deux approches différentes :

En effet et en premier lieu, nous avons développé une base de données d'images de caractères arabes manuscrits pour pouvoir expérimenter notre système. Dans le second lieu, nous avons conçus et met en place un système de reconnaissance des caractères arabe manuscrit.

Notre mémoire se subdivise donc comme suit :

Le premier chapitre sera consacré essentiellement à l'écriture arabe, nous présentons tout d'abord un aperçu général de la langue arabe et les caractéristiques de l'écriture arabe, Nous présenterons par la suite les différentes structures de document arabe

Nous présenterons dans le chapitre 2 l'architecture générale et les différentes étapes qui composent le processus de reconnaissance de l'écriture d'une façon générale. Nous exposerons ainsi les techniques les plus souvent citées et utilisées au niveau des prétraitements, d'extraction de primitives, de ségmentation, de reconnaissance.

Au chapitre 3, Nous présenterons les différentes méthodes de classification utilisée. Premièrement nous introduisons la technique des réseaux de neurones avec toutes ses notions de base. On y trouve une description de l'architecture générale d'un réseau et de son mode de fonctionnement, et, en particulier pour un perceptron et un réseau multicouches. Aussi on y présente les différentes étapes de mise en œuvre d'un réseau de neurones. Deuxièmes nous donnons une vision générale sur les SVMs, on y rappelé les principales propriétés des SVMs.

Dans le dernier chapitre du Simulations et résultats, nous présentons les résultats de reconnaissance obtenue par l'application des différentes techniques.

Nous finirons ce travail par une conclusion et des perspectives sur les travaux futurs dans ce domaine de recherche.

Notre système a donné des bons résultats au niveau de la reconnaissance avec un de taux de reconnaissance obtenu par les réseaux de neurone est 31,4% et 38,47% par les SVM et pour le matching est 88.57 %. Cela montre l'efficacité de la méthode utilisée pour l'extraction des primitives et la stratégie utilisée pour la recherche et la sélection des paramètres de modèles SVM multi-classes et réseaux de neurone.

Les perspectives envisageables pour l'amélioration et la poursuite de nos travaux sont nombreuses. Celles que nous avons jugées importantes sont les suivantes :

- Améliorer le processus de segmentation de mot en graphèmes
- Elargir la base de données de caractères manuscrits, en ajoutant d'autres exemples des ligatures extraits des documents réels.
- Utiliser autres caractéristiques qui améliorent le taux de reconnaissance

BIBLIOGRAPHIES

- [1] Abdi H, "Neural Network", Lewis -Beck M, Bryman. A, Futing T (Eds.), Encyclopedia of Social Sciences Research Methods, Thousand Oaks (CA): Sage , 2003
- [2] Abdi H, "Précis de Connexionnisme", Intelligence Artificielle et Intelligence Naturelle, Paris : Presse Universitaires de France in J. P Rossi, J. F Le Ny, J. F (Eds.), 1993
- [3] a.belaid.y. belaid. "Reconnaissance des forms methods et application » inter edition.1992
- [4] Achour K, Zenati N, Djekoune O, "Contribution d'image par un modèle de réseaux de neurones", 1998 ANN
- [5] Al badr B, mahmoud S. A. « survey and bibliography of arabic optical text recognition" signale processing, vol 41, pp:49-77,1995
- [6] Alexander C. B, Jitendra M, "Geometric Blur for Template Matching", 2001
- [7] B. Al-Badr, S.A. Mahmoud: « Survey and bibliography of Arabic optical text recognition ». Signal processing, vol. 41, pp. 49-77, 1995.
- [8] BenJ.A.krose P.patrick van der smagt, " An introduction to neural networks" university of amsterdam faculty of mathematics and computer science: pages 33- 42, Fifth edition january 1993
- [9] Bensefia A., "Analyse des documents manuscrits : identification et vérification du scripteur", Thèse de Doctorat, Laboratoire LITIS, Université de Rouen, Mai 2004
- [10] Boser, B., Guyon, I., and Vapnik, V., 1992. A training algorithm for optimal margin classifiers, In fifth annual workshop on computational learning theory, Proceedings of the 5th annual workshop on Computational learning theory, Pittsburg, pp. 144-152.
- [11] Burges, C.J.C., 1998. A Tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery Edited by Usama Fayyad, Vol. 2, pp. 121 167.
- [12] Cohen David, 1993, « Arabe : langue arabe » Enc. Universalis, vol 2, pp 707 – 714. Encyclopédie de l'Islam, 1975-, collectif , « Arabiyya » vol 1, pp579-601.
- [13] Cortes, C., and Vapnik, V., 1995. Support vector network, Machine Learning, Vol. 20, pp 273-297
- [14] Dayan P, Abbott L. F, " Theoretical Neuroscience", Book, 2000
- [15] Dreyfus G, "Les Réseaux de Neurones, Mécanique Industrielle et Matériaux, No.51, septembre 1998.
- [16] Doermann and S. Jaeger. Arabic and Chinese Handwriting Recognition, Springer, Germany, 2006.

- [17] Dr Suici-Meslati, « Reconnaissance de formes et écriture arabe manuscrite », support de cours 2008.
- [18] Dreyfus G, Martinez J-M, Samuelides M, Gordon M. B, "Réseaux de Neurones Méthodologies et Application", S. Thiria, L Herault (Eds.), Groupe Eyrolles, ISBN: 2- 212 11464-8, 2004.
- [19] El-Yacoubi M. A., Gilloux M., Bertille J. M., "A statistical approach for phrase location and recognition within a text line: an application to street name recognition", IEEE Transactions on PAMI, Vol.24, N°2, pp.172-188, February, 2002.
- [20] E.Augustin. Reconnaissance de mots manuscrits par systèmes hybrides Réseaux de Neurones et Modèles de Markov Cachés, these de doctorat, Université RENE DESCARTES - PARIS V,2001.
- [21] Fausett L, "fundamentals of Neural Networks Architectures, Algorithms and Applications", Addison Wesley, Publishing Company, Inc, 1991
- [22] Freeman J. A, Shapriua D.M, "Neural Networks: algorithms, Application and Programming Techniques", Book, Addison Wesley Publishing Company, INC, 1991.
- [23] F. Menasri, "Contributions à la reconnaissance de l'écriture arabe manuscrite," in thèse de l'Université Paris Descartes, 2008.
- [24] Jain A. K, Mao J, Muhiuddin K. M, "Artificial Neural Networks: A tutorial", IEEE, pp.31-44, March 1996
- [25] j.anigbogu « reconnaissance de textes imprimés mutifonts à l'aide de modeles stochastiques et métrique » these de doctorat,université de nancy1,1992.
- [26] J.Revaud, G. Lavoué, and A. Baskurt. Une nouvelle mesure de distance entre descripteurs de moments de zernike pour une similarité optimale et un angle de rotation entre les images. In CORESA, March 2009.
- [27] Haitaamar Schahrazed, « Segmentation de textes en caractères pour la reconnaissance optique de l'écriture arabe ». Thèse de magister université Elhadj-Lakhdar Batna, 2007.
- [28] Héroux P., Diana S., Ribert A., Trupin E., "Etude de méthodes de classification pour l'identification automatique de classes de formulaires", Actes CIFED'98, 5ème Colloque International Francophone sur l'Ecrit et le Document, pp.463- 472, Québec, Canada, Mai 1998
- [29] G. Badrinath, Naresh Kachhi, and Phalguni Gupta. Verification system robust to occlusion using low-order zernike moments of palmprint subimages. Telecommunication Systems, 47 :275-290, 2011.

- [30] K Kuhnke, K., Simoncini, L., and Kovács-V, Zs.M., 1995. A System for machine written and hand-written character distinction, International Conference on Document Analysis and Recognition ICDAR'95, Montreal, Canada, Vol. 2, pp. 811-814
- [30] K. Jain, R. P.W. Duin and J. Mao, "Statistical pattern recognition: A review", IEEE Trans, on PAMI, vol.22 (1), pp.4-37, 2000.
- [31] L. CHERGUI « Combinaison de classifieurs pour la reconnaissance de mots arabes manuscrits » Thèse de doctorat . Université Mentouri Constantine.2013 lailaa chapitre 4
- [32] LeBourgeois F., Emptoz H,. "DEBORA: Digital accEs to BOoks of the RenaissAnce", International Journal on Document Analysis and Recognition, IJDAR , Vol. 9, N°. 2, pp.193-221, April 2007.
- [33] M. Cheriet, N. Kharm, C.L. Liu, and C.Y. Suen. Character Recognition Systems: A Guide for Students and Practitioners, John Wiley & Sons, New Jersey-United States of America (USA), 2007.
- [34] M. Mansour, M. Benkhadda & A. Benyettou, « Optimized Segmentation Techniques 90 for Handwritten Arabic Word and Numbers Character Recognition », 2005 Segmentation
- [35] M. Sarfraz. Computer-Aided Intelligent Recognition: Techniques and Applications, Wiley, England, 2005. .
- [36] M. R. Teague. Image analysis via the general theory of moments □. J. Opt. Soc. Am., 70(8) :920-930, Aug 1980
- [37] Marc parizeau, "Réseaux de neurones", édition 2004
- [38] Michel Neyreneuf et Ghalib Al-Hakkak, Grammaire active de l'arabe, Le Livre de Poche, collection « Les langues modernes », 1996
- [39] Muhammad Sarfraz, « Computer -Aided intelligent recognition techniques and applications » (section: Offline Arabic Character Recognition) King Fahd University of Petroleum and Minerals, Kingdom of Saudi Arabia, 2005
- [40] N. Mezghani, A. Mitiche, et M. Cheriet, "A new representation of shape and its use for high performance in online Arabic character recognition by an associative memory," International Journal on Document Analysis and Recognition, vol. 7, n°. 4, p. 201-210, 2005.
- [41] Najoua Ben Amara et al, « Modélisation Pseudo Bidimensionnelle pour la Reconnaissance de Chaînes de Caractères Arabes Imprimés », 1999.
- [42] N Nemmour, H., 2010. Application des machines à vecteurs de support pour la reconnaissance des chiffres manuscrits. Thèse de doctorat, USTHB Alger, Algérie, p 171.

- [43] Philippe BOLON ET AL, « Analyse d'image, Traitements de bas niveau » (support de cours), 1998).
- [44] Principe J.C. "Artificial Neural Networks", The Electrical Engineering Handbook Eds. Richard C. Dorf Boca Raton: CRC Press LLC, 2000.
- [45] Redouane TLEMSANI , 'LA RECONNAISSANCE EN-LIGNE DU MANUSCRIT ARABE' Thèse de DOCTORAT , UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE D'ORAN -MOHAMED BOUDIAF , 2012
- [46] R. Swiniarski. Rough sets and neural networks application to handwritten character recognition by complex zernike moments. In Lech Polkowski and Andrzej Skowron, editors, Rough Sets and Current Trends in Computing, volume 1424 of Lecture Notes in Computer Science, pages 617–624. Springer Berlin / Heidelberg, 1998.
- [47] S. Chevalier. Reconnaissance d'écriture manuscrite pour des techniques markoviennes: une approche bidimensionnelle et générique, Thèse de Doctorat, Université René Descartes, France, 2004.
- [48] S. Kermi : « Classifieur neuronal base connaissances, application a la reconnaissance des caractères arabes isoles manuscrits ». Thèse de magister, université Badji Mokhtar, Annaba, Algérie 1999.
- [49] S. Snoussi maddouri, « Modèle perceptif neuronal à vision globale-locale pour la reconnaissance de mots manuscrits arabes », 2002.
- [50] SAADY, ' Contribution au développement d'approches de reconnaissance automatique de caractères imprimés et manuscrits, de textes et de documents Amazighes ' THÈSE DE DOCTORAT EN SCIENCES Université Ibn Zohr Faculté des Sciences d'Agadir , 2012
- [51] Sander Bohte, « Spiking neural networks », Thesis University Leiden.2003,
- [52] Shih-chi huang and yih-fang huang, "Learning algorithms for perceptrons using back-propagation with selective update" paper, 1990
- [53] Simon haykin'' Neural networks a comprehensive foundation'' ISBN0-13- 908385-5 prentice hall international, Inc pages: 34-78, second edition, 1999
- [54] Ss
- [55] S.N. Srihari & E.J. Keubert. Integration of handwritten address interpretation technology into the United States postal service remote computer reader system. ICDAR, pp.892–896, 1997.
- [56] Sylvain Chevalier, « Reconnaissance d'écriture manuscrite par des techniques markoviennes: une approche bidimensionnelle et générique ». Thèse de doctorat de l'Université René Descartes - Paris 5, 2004.

[57] Trevor Hastie and Patrice Y. Simard, « Metrics and Models for Handwritten Character Recognition ». Statistical science Vol. 13 N° 1, 54 -65, 1998.

[58] V. Venkatraman, P. Chakravarthy, and D. Kihara. Application of 3d zernike descriptors to shape-based ligand similarity searching . Journal of Cheminformatics, 1(1) :19, 2009.

[59] mathworks , <http://es.mathworks.com/> , consulté le : 15/05/2015.

[60] Yousef Al-Ohali, MohamedCheriet et Ching Suen, « Databases for recognition of handwritten Arabic cheques ». Pattern Recognition Society 36 111 – 121. 2003.

[61] ZEMOURI ET-Tahir , ‘ Discrimination Texte Manuscrit – Texte Imprimé par les SVMs’

ملخص:

تعالج هذه المذكرة عملية إنشاء نظام جديد للتعرف على الحروف العربية المكتوبة باليد، والذي يتمحور حول تطبيق مختلف المصنفات منها الشبكات العصبية والآلات الحاملة للأشعة، وكذلك إنشاء قاعدة بيانات جديدة للحروف العربية المكتوبة. مجموعة من العمليات المخصصة للمعالجة تطبق على الصور، الهدف منها تقليص سمك الخط، توحيد الحجم، تحويل الصورة إلى لغة مفهومة لدى الحاسوب بواسطة الماسح الضوئي، مع ذلك فإن نقل المعلومات يؤدي إلى فقدان البعض منها، لذلك قمنا في عملنا هذا بتطوير وتقييم نظام جديد للتعرف على الحروف العربية المكتوبة بخط اليد على مختلف وحدات التصنيف. الكلمات المفتاحية: التعرف على الحروف لعربية، الشبكات العصبية، الآلات الحاملة للأشعة

Abstract:

This memory is mainly dedicated to the development of a new powerful offline recognition system of Arabic handwritten characters based on different classifiers which are: ANNs and SVMs. And to the design of a new database of Arabic characters, some preprocessing operations are applied to images' characters, including likes mousing, binarization, skew and normalization, To extract primitives used to classify words, geometrical moments are used: ZERNIKE moments; the last was exploited by a distinct different classifiers. The evaluation of the proposed recognition system is done on our new database. The proposed system reports good results but it can be comparable to the best systems of Arabic handwriting recognition presents in the literature.

Keywords: Arabic handwriting recognition, combined classifiers, neural networks, geometrical moments, handwritten Arabic words databases.

Résumé:

Le travail présenté dans ce mémoire est consacré le développement d'un système de reconnaissance hors-ligne des caractères arabes manuscrits basé sur des différents classifieurs qui sont : les ANNs, et SVMs. Et à la conception d'une nouvelle base de données des caractères arabes, une série d'opérations de prétraitement est appliquée aux images de caractères, incluant : un lissage, une normalisation, and correction d'inclusion et binarisation, pour l'extraction des caractéristiques des caractères on trouve les moments de ZERNIKE, ce dernier a été exploité par les différents classifieurs, L'évaluation du système de reconnaissance proposé est menée sur la nouvelle base que nous avons conçue. Les résultats montrent que le système proposé donne des bons résultats mais elles sont comparables à ceux des meilleurs systèmes de reconnaissance de caractères arabes manuscrits présents dans la littérature.

Mots-clés : Reconnaissance de l'écriture arabe manuscrite, classifieurs, Réseaux de neurones, SVM, Moments géométriques, Base de caractères arabes.