HAL
open science

Une approche d'identification des fontes arabes

Najoua Essoukri Ben Amara, Sami Gazzah

To cite this version:

Najoua Essoukri Ben Amara, Sami Gazzah. Une approche d'identification des fontes arabes. Jun 2004. sic_00001173

HAL Id: sic_00001173

https://archivesic.ccsd.cnrs.fr/sic_00001173

Submitted on 6 Dec 2004

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.

Une approche d'identification des fontes arabes

Najoua Essoukri Ben Amara[1]— Sami Gazzah[2]

[1]Ecole Nationale d'Ingénieurs de Monastir - 5019 Monastir – TUNISIE
Laboratoire de Système et de Traitement de Signal-ENIT

[1]Najoua.BenAmara@enim.rnu.tn, [2]cos@email.ati.tn

Résumé : *Le développement des outils d'édition a conduit à des documents composites constitués de différents types d'informations (texte, graphique, formules, tableaux…) ayant une mise en page de plus en plus complexe. Dans les entités physiques textes, l'utilisation de plusieurs fontes à la fois dans différents attributs (gras, italique, souligné…) est devenue très courante. Dans ce contexte, connaître l'identité de la fonte devient indispensable pour une meilleure analyse des documents et de meilleures performances de reconnaissance. Ce travail traite des difficultés liées à l'identification de la fonte dans un contexte de lecture optique de l'Arabe multifonte. Ces difficultés découlent principalement de la complexité morphologique de cette écriture et de la grande variabilité, d'une fonte à une autre, des dessins des différents caractères. Nous proposons par ailleurs, une approche d'identification des fontes arabes, basée sur l'exploration des ondelettes en paramétrisation et les réseaux de neurones en classification. Les différents tests effectués sur un jeu de 114229 pseudo-mots (chaînes de caractères) dans neuf fontes, considérées dans cinq corps différents, ont conduit à des taux de reconnaissance très encourageants.*

Mots-clés : OFR, fonte arabe, approche a priori, OCR, ondelettes, réseaux de neurones, multi-corps.

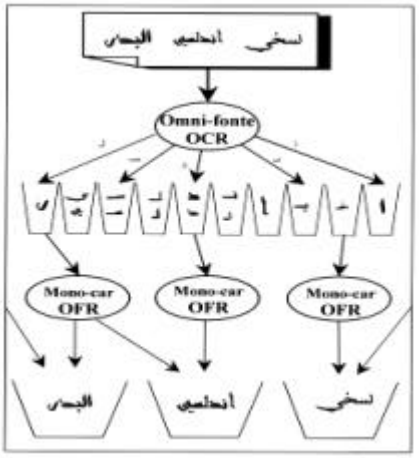
1 Introduction

La reconnaissance de l'imprimé peut être mono-fonte, multifonte ou omni-fonte. Un système OCR-Optical Character Recognition est dit mono-fonte s'il ne traite qu'une seule fonte à la fois, c'est à dire qu'il ne connaît que le graphisme d'une fonte unique. Ce cas est simple particulièrement au niveau de l'apprentissage car l'alphabet représenté est réduit. Un système OCR est dit multifonte s'il est capable de reconnaître divers types de fontes parmi un ensemble de fontes préalablement apprises. Dans ce cas, le traitement doit réduire les écarts entre même caractères (taille, épaisseur et inclinaison). Un système OCR omni-fonte est capable de reconnaître toute fonte, généralement sans apprentissage, en se basant sur les règles topologiques et morphologiques de l'écriture.

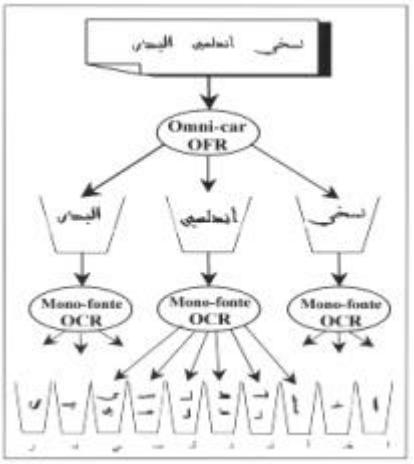
Malgré l'importance de l'information issue de l'identité de la fonte dans un contexte OCR, la reconnaissance optique de la fonte (OFR-Optical Font Recognition) reste un problème souvent négligé et les études dans ce domaine sont peu nombreuses notamment dans le cas de l'Arabe. Certains auteurs se contentent de l'identification de la famille de la fonte [Shi 97].
L'OFR est avantageuse et nécessaire surtout dans les domaines suivants [Zramdini 92] :

- La reconnaissance de la structure des documents, où la reconnaissance de la fonte du mot, de la ligne de texte ou encore d'un bloc de texte peut être utile pour la définition de ses entités logiques (le titre du document, ceux des sections ou des paragraphes).
- La reproduction du document, pour laquelle la connaissance de la fonte est nécessaire dans la reconstitution du document dans sa version originale.
- L'indexation du document et l'extraction d'information, les indexes mot étant généralement imprimés dans des fontes différentes de celle du texte courant.
- La reconnaissance optique des caractères, dans ce cas l'OFR permet une amélioration considérable des performances de reconnaissance.

Il existe principalement deux stratégies de combinaison de l'OFR avec l'OCR : l'approche a posteriori et l'approche a priori [Zramdini 95]. L'identification a posteriori (voir Fig. 1-a), reconnaît la fonte après avoir reconnu les caractères. Elle aide à corriger les erreurs de reconnaissance et permet de retrouver la description d'origine du document. L'identification a priori (voir Fig. 1-b) reconnaît la fonte sans pour autant reconnaître les caractères. Elle simplifie considérablement les tâches ultérieures dans un système OCR en optimisant les recherches par la réduction du nombre de graphismes différents à considérer. De plus, elle améliore les performances globales du système en bénéficiant des différentes informations relatives à la fonte, préalablement disponibles.
La Figure 1 illustre le principe des deux approches de l'OFR, une méthode de type analytique est supposée être utilisée.



(a)



(b)

FIG 1- stratégies de combinaison de l'OFR avec l'OCR, (a) a posteriori, (b) a priori, [Gazzah 03a].

Dans ce contexte, nous proposons une contribution à l'étude et à la caractérisation des fontes Arabes. Ce travail s'insère dans un cadre plus général d'analyse et de reconnaissance de documents arabes. Les différents travaux développés au sein de notre équipe nous ont montré l'utilité et l'importance de l'utilisation de sources d'informations multiples parmi lesquelles se trouve l'identité de la fonte dans le cas de l'imprimé [Essoukri Ben Amara 03]. D'où l'idée de développer un classifieur de fontes arabes en vu de l'intégrer dans un système AOCR-Arabic OCR, et de se ramener à des problématiques de reconnaissance mono-fonte.

Dans la section suivante nous proposons une étude sur les fontes arabes à travers laquelle nous montrons la complexité de l'AOCR imprimé multifonte. Dans la section 3, un état est présenté sur le multifonte en AOCR, nous verrons que cet axe reste très peu exploré. Nous proposons dans la section 4 une approche d'identification des fontes arabes basée sur l'utilisation des ondelettes en paramétrisation et les réseaux de neurones en classification. Les différentes expérimentations ainsi que les résultats enregistrés sont détaillés dans la section 5.

2 Typologie des fontes Arabes

L'écriture Arabe est connue pour sa richesse en fontes et styles. Il existe environ 450 fontes d'écritures différentes dont seulement quelques-unes sont couramment utilisées dans le monde arabo-musulman, nous citons à titre d'exemple : le Neskhi, Thoulthi, Roqa, Diwani, Koufi, Farsi… Le Neskhi demeure aujourd'hui la fonte la plus utilisée pour l'écriture imprimée. Chaque style arabe est régi par des lois particulières. D'un style à un autre, les proportions d'une même lettre et son dessin peuvent changer considérablement.

Nous avons mené une étude sur un nombre important de fontes Arabes [Gazzah 04a]. L'examen des dessins des caractères montre des différences considérables inter et intra-fontes à différents degrés. Parmi les variations les plus importantes nous retenons les différences aux niveaux :

- du dessin et/ou du nombre de boucle d'un caractère (voir Fig. 2-a),
- des dimensions ainsi que la forme aussi bien des extensions hautes (hampe) que basses (jambage) (voir Fig. 2-b),
- du dessin des points diacritiques (voir Fig. 2-c) et ceux des signes diacritiques tels que Shadda " ؟ " et Madda "~" (voir Fig. 2-b, 2-c), ainsi que leur position par rapport au tracé du caractère,
- des accolements entre pseudo-mots (voir Fig. 2-d).
- de la disposition spatiale du caractère par rapport à la ligne d'écriture (naturellement horizontale) qui varie dans certaines fontes (voir Fig. 2-e).

Par ailleurs, certains caractères, contrairement aux règles morphologiques de l'Arabe, peuvent garder le même dessin dans des positions différentes (début/milieu, début/fin) (voir Fig. 2-f). En outre, la disposition de la ligne d'écriture varie en fonction de l'identité de la fonte (voir Fig. 2-g). Dans certaines fontes plusieurs caractères peuvent être écrits de façon combinée. Ces combinaisons ou ligatures verticales, dont le nombre dépasse 150, sont optionnelles [Haralambous 95]. Elles sont utilisées généralement pour des raisons d'esthétique et/ou d'édition de texte. Les ligatures verticales dépendent du dessin de la police et du degré de qualité artistique du document. Elles peuvent être formées de deux, trois ou quatre caractères. Les accolements verticaux entre pseudo-mots (chaînes de caractères) d'un même mot différent d'une fonte à une autre (voir Fig. 2-h).

3 Le multifonte en AOCR

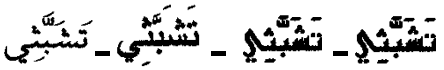
Pour toutes les caractéristiques morphologiques citées précédemment (et pour bien d'autres), l'identification de la fonte dans le cas de l'Arabe est une tâche non triviale, cependant elle est indispensable afin de réduire un tant soit peu la complexité de l'AOCR. Peu de travaux en reconnaissance de l'Arabe utilisent des contextes multifonte (en moyenne 2 à trois fontes) sans pour autant procéder à une étude de la fonte du

document considéré [Abdelazim 89, Aissaoui 96, Amin 86, Amin 89, Amin 97, Bushofa 97, Fehri 94, Kurdy 93]. D'après nos références, seuls les travaux suivants [Ben Amara 96, Ben Mahmoud 98, Gadhoumi 97] ont étudié le problème de la reconnaissance de la fonte.

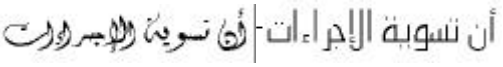
Dans les travaux de Gadoumi *et al*. [Gadoumi 97] un système de classification neuronal de fontes arabes a été proposé. Les fontes Jiza, Nedim, Andalus, Andalus bold, Transparent, Mudir, Mudir bold, Simplified, Boutress, Koufi, ont été considérées. Vingt six paramètres ont été retenus pour la caractérisation des différentes fontes. Sept paramètres sont extraits de la ligne de référence, six de la première dérivée du profil vertical et de la densité des mots en pixel. Les treize paramètres restants sont des écarts types statistiques sur les valeurs des treize premiers paramètres. Les expérimentations ont montré que l'intégration de l'AOFR dans le système AOCR développé a conduit à une amélioration des performances globales de plus de 3%. Dans [Ben Amara 96], un module d'identification de la fonte est proposé. La méthode se base sur l'extraction de certaines caractéristiques structurelles de l'écriture telles que l'épaisseur horizontale et verticale du trait, la largeur de la bande centrale, la densité de pixels noirs d'un point diacritique, la largeur moyenne d'un caractère isolé. Cinq fontes ont été considérées : Kahra, Nedim, Iskandaria, Mehdi et Jiza. Par ailleurs, la modélisation retenue (planaire de type markovien) des pseudo-mots Arabes, est indépendante des variations du corps de la fonte.

(a) Exemples de dessins de boucle dans des fontes différentes.

(b) Exemples de dessins de hampe et de jambage dans des fontes différentes.



(c) Variation du dessin et de la position des diacritiques en fonction de la fonte.

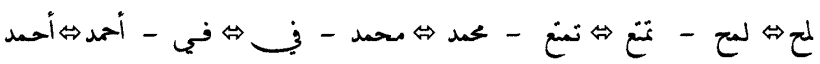


(d) Différents accolements de pseudo mots dans des fontes différentes.

(e) Disposition des caractères par rapport à la ligne d'écriture.

(f) Exemples de dessins de caractères pour une même position contextuelle.

(g) Variation de la ligne d'écriture en fonction de la fonte.



(h) Pseudo-mots avec et sans ligatures (à gauche et à droite du « ⇔ » resp.).

FIG 2- Illustrations de certaines variations morphologiques de l'Arabe en fonction de l'identité de la fonte.

4 Un OFR pour un système AOCR multifonte

Il est vrai que l'écriture imprimée présente des caractères bien alignés avec un graphisme conforme à un style calligraphique, cependant ceci n'est pas tout à fait le cas de l'Arabe étant donné les multiples variations morphologiques inter et intra-fontes. L'identification de la fonte s'avère donc indispensable afin de réduire la complexité de l'AOCR.

Dans ce contexte nous avons développé un classifieur de fontes Arabes en vu de l'intégrer dans des systèmes AOCR multifonte. L'approche OFR adoptée est de type a priori, la reconnaissance de la fonte s'opère en amont de l'OCR. Nous avons opté pour les ondelettes au niveau de la phase de paramétrisation et les réseaux de neurones au niveau de la phase de classification. La Figure 3 situe le module OFR développé dans un contexte global de reconnaissance d'un texte arabe multifonte.

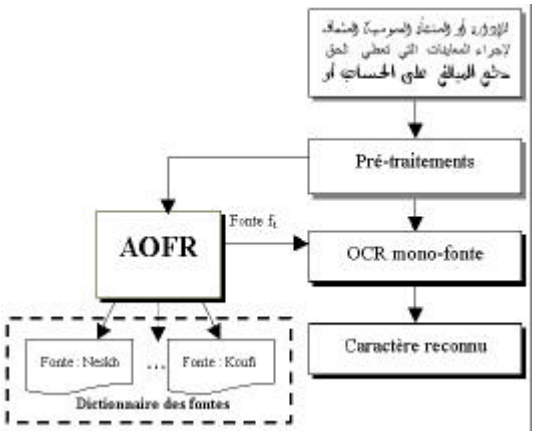


FIG. 3. Schéma global d'un système AOCR multifonte doté d'un module OFR en amont.

4.1 Système AOFR développé

Les images numérisées des différents textes sont d'abord segmentées en lignes de texte et en pseudo-mots (ou plusieurs pseudo-mots dans le cas d'accolements). Les images des pseudo-mots sont alors filtrées puis normalisées en hauteur afin de s'affranchir des problèmes de variation du corps de la fonte [Chen 99]. L'extraction des primitives est assurée par la technique des ondelettes, la classification est de type neuronale. Nous nous limitons dans ce qui suit aux étapes d'extraction des primitives et de classification.

4.2 Extraction des primitives

Nous avons expérimenté plusieurs types de primitives (notamment statistiques, et structurelles). Puisque nous considérons les pseudo-mots sans segmentation au préalable en caractères (ou en graphèmes) et à cause de la grande variabilité morphologique inter et intra-fontes, des primitives globales sont avérées mieux appropriées. Les ondelettes, fréquemment utilisées en compression d'images, ont été expérimentées. Ce sont des transformations qui permettent à l'image de passer à une autre représentation produisant des vecteurs dont les composantes sont de variances différentes [Akansu 92]. Nos travaux en compression d'images [Ben Amor 02] nous ont incités à les expérimenter dans le cas de l'AOCR. Cette technique commence à être implémentée en reconnaissance de l'écriture [Akansu 92, Bhattacharya 02, Pujari 02,Yang 03]. Les différents tests effectués nous ont conduit à l'utilisation des ondelettes de Daubechies [Gazzah 03a, Gazzah 03b]. Une décomposition à l'ordre 3 a été adoptée, un ordre supérieur produit une perte importante d'information. En plus des paramètres issus des ondelettes, nous avons considéré la densité en pixels noirs des pseudo-mots, en effet, les différents tests ont montré l'apport de cette primitive dans la caractérisation. Les primitives retenues sont alors les suivantes : la densité en pixels noirs, l'écart type et la moyenne des 4 matrices issues des ondelettes correspondant respectivement à l'image d'approximation, au détail horizontal, au détail vertical et au détail diagonal.

5 Classification

Il s'agit de déterminer l'identité de la fonte étant donné le jeu de primitives préalablement extrait. A ce niveau, nous avons opté pour les réseaux de neurones multicouches. Nous avons expérimenté deux types de réseaux: les perceptrons multicouches-PMC et les réseaux à fonction radiale de base-RBF. Pour chacun de ces réseaux, deux architectures ont été testées : un réseau global à « n » sorties et un réseau par fonte à une seule sortie. Les meilleurs résultats ont été enregistrés dans le cas d'un réseau de type RBF modulaire ; nous nous limitons dans ce travail aux expérimentations relatives à ce réseau, les résultats du réseau PMC ont été publiés ultérieurement dans [Gazzah 04b].

6 Expérimentations et résultats

Les différentes expérimentations ont été effectuées sur un jeu de 114229 pseudo-mots (86516 pseudo mots pour l'apprentissage et 27713 pour les tests) issus de textes extraits du Journal Officiel de la République Tunisienne. Les différents textes ont été considérés dans neuf fontes parmi les plus communément utilisées dans les applications tournant sous Windows (95, 98, 2000). Il s'agit des fontes : Arabic transparent, Badr, Andalous, Alhada, Diwani, Naskhi, Koufi, Times New Roman et le Thouloth. Nous avons considéré ces fontes dans les corps : 12, 18, 24, 28 et 36. Les fontes Arabic Transparent (AT) et Times New Roman (TNR) présentent une très forte ressemblance morphologique ; l'étude montre seule une différence au niveau des élongations horizontales inter-caractères. Ce choix nous a permis d'évaluer les performances de notre approche dans le cas de fontes similaires. Les différentes expérimentations ont montré des taux de reconnaissance très encourageants pour la majorité des fontes (cf. Tab. 1). A cause de la très grande similarité morphologique qui existe entre les fontes AT et TNR, les taux de reconnaissance de ces fontes sont relativement faibles. Nous avons reconduit les tests en considérant ces deux fontes comme étant une seule (AT&TNR), les résultats ont confirmé cette mesure, en effet, une nette amélioration des taux de reconnaissance a été enregistrée (cf. Tab. 2).

Fonte	Alhada	Andalous	AT	Badr	Diwani	Koufi	Nask	TNR	Thouloth	Taux global
Taux	99.85 %	99.70 %	91.78 %	100 %	99.85 %	99.55 %	99.40 %	98.65 %	99.71%	98.72%

TAB. 1- Taux de reconnaissance par fonte,
cas où les fontes AT et TNR sont séparément considérées.

Fonte	Alhada	Andalous	Badr	Diwani	Koufi	Nask	AT&TNR	Thouloth	Taux global
Taux	99.85 %	99.70 %	100 %	99.85 %	99.55 %	99.40 %	99.24 %	99.71 %	99.67%

TAB. 2- Taux de reconnaissance par fonte, avec la fonte AT&TNR.

5 Conclusion

Dans ce travail, nous avons considéré la problématique liée à l'identification des fontes arabes dans un contexte de reconnaissance de l'écriture arabe multifonte, multi-corps. Une étude sur la morphologie de l'Arabe a montré que l'AOFR est une tâche non triviale, cependant elle est nécessaire afin de réduire la complexité de l'AOCR et améliorer les performances de reconnaissance. Nous avons par la suite, proposé une approche d'identification de fontes arabes basée sur l'exploration des ondelettes en caractérisation et des réseaux de neurones de type RBF en classification. Deux architectures de réseaux RBF ont été expérimentées : un réseau global à sorties multiples et un réseau par fonte, à une seule sortie. Il a été trouvé que le modèle à une seule sortie est plus performant que celui à plusieurs sorties (99.45 % vs 99.11 %). Pour les deux fontes AT et TNR, qui présentent une très forte ressemblance morphologique, nous avons enregistré des limitations dans leur discrimination. La considération des deux fontes comme étant une seule, a conduit à une amélioration considérable des performances des deux réseaux. Nous poursuivons les expérimentations afin d'étudier plus profondément le problème de la similarité entre fontes ainsi que les cas de rejet.

6 Bibliographie

[Abdelazim 89] ABDELAZIM H.Y., HASHISH M.A.: "Interactive font learning for Arabic OCR". *Proc. 1st Kuwait* computer conference, pp. 463-486, Kuwait, 1989.

[Aissaoui 96] AISSAOUI H., AISSAOUI A., HAOUARI A.: "Une approche neuronale pour la reconnaissance de texte arabe imprimé multifonte ». *16èmes Journées tunisiennes en électrotechnique et automatique, JTEA'96*, pp. 402-409, Hammamet-Nabeul, Tunisie, 1996.

[Akansu 92] A. N. AKANSU and R. A. HADDAD:" Multiresolution Signal Decomposition : Transform, Subbands and Wavelet's". *Academic Press, Inc.* (1992).

[Amin 86] A. AMIN, G. MASINI : "Machine recognition of multi-font printed arabic texts". *Proc. 8th Int. Joint Conf. on Pat. Recognition*, pp. 392-395, France, 1986.

[Amin 89] A. AMIN, J.F. MARI : "Machine recognition and correction of printed Arabic text". *IEEE Transactions on systems, man, and cybernetics*, vol.19, n 5, pp. 1300-1304, September/October 1989.

[Amin 97] A. AMIN, W. MANSOOR : "Recognition of printed arabic text using neural networks". *IEEE Proc. 4th Int. conf. on document analysis and recognition, ICDAR'97*, pp. 612-615, Ulm, Germany, 1997.

[Anigbogu 92] J. ANIGBOGU : "Reconnaissance de textes imprimés multifonte à l'aide de modèles stochastiques et métriques". Thèse de doctorat, Université de Nancy I, 1992.

[Ben Amara 96] N. BEN AMARA : "PHMM pour PAWs imprimés arabes : vers une indépendance du corps de la fonte". *16èmes Journées tunisiennes en électrotechnique et automatique, JTEA'96*, pp.330-339, Hammamet-Nabeul, Tunisie, 1996.

[Ben Amor 02] N. BEN AMOR, N. ESSOUKRI BEN AMARA : «DICOM Image Compression By Wavelet Transform ». *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*, vol.2 6-9 Octobre 2002.

[Ben Mahmoud 98] N. BEN MAHMOUD, M.C FEHRI, H. AMIRI, K. GADHOUMI, H. HAOUALA : "Radial bases function network for arabic font recognition". *Computational engineering in systems applications, CESA'98*, pp.13-16, Nabeul-Hammamet, Tunisie, 1998.

[Bhattacharya 02] U. BHATTACHARYA, T.K. DAS, CHAUDHURI : " A Cascaded Scheme for Recognition of Handprinted Numerals'', *Indian Conference on Computer Vision, Graphics and Image Processing : ICVGIP 2002.*

[Bushofa 97] B.M.F. BUSHOFA : "Computer recognition of printed arabic characters". Ph.D Thesis, Faculty of Engineering of the university of Birmingham, 1997.

[Chen 99] G. CHEN : "Application of wavelet transfoms in pattern recognition and de-noising". Ph.D These, Concordia University, Canada 1999.

[Essoukri Ben Amara 99] N. ESSOUKRI BEN AMARA : "Utilisation des modèles de Markov cachés planaires en reconnaissance de l'écriture arabe imprimée". Thèse de doctorat, Université Tunis II, 1999.

[Essoukri Ben Amara 03] N. ESSOUKRI BEN AMARA F. BOUSLAMA: "Classification of Arabic Script Using Multiple Sources of Information: State of the Art and Perspectives". "*Special issue on Multiple Classifiers, International Journal on Document Analysis and Recognition (JDAR 2003)*, Vol. 5, Number 4, July 2003.

[Fehri 94] M.C. FEHRI, M. BEN AHMED : "A new approach to arabic character recognition in multifont documents". *Proc. 4th Int. conf. and exhibition on multi-lingual computing (Arabic and Roman Script)*, pp. 2.5.1-2.5.7, April 1994.

[Gadhoumi 97] K. GADHOUMI, N. BEN MAHMOUD : "Reconnaissance des fontes arabes par réseaux de neurones à base de fonctions radiales". Projet de fin d'études, Ecole Nationale d'Ingénieurs de Tunis, Juin 1997.

[Gazzah 03a] S. GAZZAH, N. ESSOUKRI BEN AMARA : « Problématique liée à l'identification des fontes arabes ». *Proc. Journées Scientifique en Génie Electrique et Informatique, GEI 2003*, vol. 1, pp. 122-127, 17-19 mars Mahdia, Tunisie 2003.

[Gazzah 03b] S. GAZZAH, N. ESSOUKRI BEN AMARA , "Etude et Caractérisation des Fontes Arabes". *Int. Conf. On Image and Signal Processing, ICISP'2003*, 25-25 june, Vol. 2 pp. 650-659, Agadir, Marroc, 2003.

[Gazzah 04a] S. GAZZAH, N. ESSOUKRI BEN AMARA , " Utilisation des Ondelettes et des réseaux de neurones en identification des fontes arabes". Accepté, *Conférence Internationale Signaux, Circuits et Systèmes, SCS* 2004.

[Gazzah 04b] S. GAZZAH, : « Contribution à l'étude et la caractérisation des fontes arabes ». Mémoire de mastère, Ecole Nationale d'Ingénieurs de Sfax,Tunisie Février 2004.

[Haralambous 95] Y. HARALAMBOUS, J. PLAICE : "O, une extension de TEX incluant UNICODE et des filtres de type Lex". Cahiers GUTenberg, codage des caractères d'ASCII à UNICODE, pp. 55-79, journée Oméga, Genève, 16 mars 1995.

[Kurdy 93] M.B. KURDY, A. JOUKHADAR, A. WABBI : "Multifont arabic/latin optical character recognition system". *Actes du Colloque : Langue arabe et technologies informatiques avancées*, pp.245-256, Casablanca, 1993.

[Parramon 87] J.M.a PARRAMON : "Comment dessiner lettres, titres et logotypes". Editions Bordas, Barcelona, 1987.

[Pujari 02] A.PUJARI, C. D. NAIDU et B. C. JINAGA : "An Adaptative Character Recognizer for Telugu Scipts using Multiresolution Analysis and Associative Memory ", *Indian Conference on Computer Vision, Graphics and Image Processing : ICVGIP 2002*.

[Shi 97] H. Shi, T. Pavlidis :"Font recognition and contextual processing for more accurate text recognition". *International Conference on Document Analysis and Recognition*, ICDAR'97, pp. 39-44, Ulm-Germany, 1997.

[Yang 03] F.YANG, M.PAINDAVOINE, "Pre-filtering for pattern recognition Using Wavelet Transform and Neural Networks", *Advances in Imaging & Electron Physics, Academic Press San Diego*. Vol. 127, 2003.

[Zramdini 92] A. ZRAMDINI, A. AZOKLY, R. INGOLD : « Importance de l'identification de la fonte dans la reconnaissance structurelle de documents ». *Actes Colloque National sur l'écrit et le document, CNED'92*, pp. 233-241, Nancy, 1992.

[Zramdini 93] A. ZRAMDINI, R. INGOLD: « Optical font recognition from projection profiles ». Electronic publishing, Vol. 6(3), pp. 249-260, September 1993.

[Zramdini 95] A. ZRAMDINI : "Study of optical font recognition based on global typographical features". Thèse n°1106, Université de Fribourg, 1995.