



Modèle perceptif neuronal à vision globale-locale pour la reconnaissance de mots manuscrits arabes.

Samia Snoussi-Maddouri, Abdel Belaïd, Christophe Choisy, Hamidi Amiri

► To cite this version:

Samia Snoussi-Maddouri, Abdel Belaïd, Christophe Choisy, Hamidi Amiri. Modèle perceptif neuronal à vision globale-locale pour la reconnaissance de mots manuscrits arabes.. Colloque International Francophone sur l'Écrit et le Document - CIFED'2002, 2002, Hammamet, Tunisie, 15 p, 2002. <inria-00107579>

HAL Id: inria-00107579

<https://hal.inria.fr/inria-00107579>

Submitted on 19 Oct 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.

Modèle perceptif neuronal à vision globale-locale pour la reconnaissance de mots manuscrits arabes

S. SNOUSSI MADDOURI¹, A. BELAÏD², Ch. CHOISY² et H. AMIRI¹

¹Laboratoire des Systèmes et Traitement du Signal (LSTS), Ecole Nationale d'Ingénieurs de Tunis (ENIT), B.P 37 Belvédère 1002, Tunisie

²Group READ, LORIA-CNRS Campus scientifique B.P. 239 54506 Vandoeuvre-Lès-Nancy France

samia.maddouri@enit.rnu.tn, Abdel.Belaid@loria.fr, Christophe.Choisy@loria.fr, hamid.amiri@enit.rnu.tn

Résumé

Nous proposons dans cet article un système de reconnaissance de mots manuscrit arabes basé sur les modèles perceptifs d'activation interactive et de vérification, définis par des psychologues. Le système proposé est basé sur un Réseau de Neurones faisant partie des systèmes Transparent (RNT). Il procède par une vision globale des caractéristiques structurelles apparentes durant la première phase de propagation et une vision locale par des Descripteurs de Fourier normalisés en phase de rétro-propagation (DF). L'avantage du système appelé RNT-DF, est qu'il ne nécessite pas d'étape d'apprentissage. Il se réfère à des descriptions standard de l'écriture imprimée aussi bien au niveau global (mot) qu'au niveau de la normalisation locale (lettres). Une évaluation expérimentale des deux visions est effectuée sur une base de données des montants littéraux de chèques et de noms de villes tunisiennes.

Mots clés : modèle perceptif, réseau de neurones transparent, vision globale, normalisation locale, descripteurs de Fourier.

1. Introduction

Parmi les études actuelles sur la reconnaissance de l'écriture manuscrite, certaines cherchent à copier la vision humaine dont les performances sont naturellement les plus élevées. Cependant, la méconnaissance du mécanisme de lecture humaine rend difficile sa réalisation informatique. Des expériences ont été effectuées par des psychologues tels que McClelland et Rumelhart

[McClelland 85] dans le but d'étudier dans un premier lieu la manière dont l'humain reconnaît les formes, en second lieu les caractéristiques qui stimulent le plus l'œil humain. Des expériences psycho-cognitives ont permis d'observer le comportement de l'être humain au moment de la lecture, menant aux observations suivantes :

- 1- **Importance du contexte lexical** : La vision globale peut aider à déduire une information locale présentant dans certains cas une distorsion;
- 2- **Caractéristiques flagrantes** : La vision globale peut être suffisante pour la reconnaissance d'une forme;
- 3- **Analyse fine** : En présence de formes proches, des informations additionnelles sont nécessaires;
- 4- **Prototypage des formes** : Afin de reconnaître des formes présentant des distorsions, il n'est pas nécessaire d'apprendre toutes les distorsions possibles. L'apprentissage d'un prototype typique peut être suffisant.

Sur ces principes, des psychologues ont proposé des modèles de perception de l'écriture par des types particuliers de réseaux de neurones, qui ont été implémentés par des chercheurs en lecture automatique.

Dans cet article, nous présentons une implémentation de la perception humaine de l'écriture qui s'inspire de ces travaux. Un Réseau de Neurones Transparent (RNT) traite l'aspect de vision globale du mot, et la vision locale est assurée par des Descripteurs de Fourier normalisés (DF). Ce modèle sera appelé par la suite RNT-DF. La section 2 revient sur les modèles perceptifs proposés par les psychologues et leurs implémentations informatiques. La section 3 décrit plus complètement le principe de fonctionnement du modèle RNT-DF. La section 4 aborde le problème de la normalisation locale dans le but d'améliorer l'identification des lettres. La section 5 décrit les expérimentations du RNT-DF sur la base de données IFN/ENIT.

2. Perception humaine de l'écriture

2.1. Modèles formels perceptifs

Parmi les approches proposées par les psychologues pour modéliser la lecture humaine, nous distinguons :

- Le **modèle de vérification** des stimulus visuels sur les mots qui ont été activés par ces derniers afin de trouver le meilleur candidat [Becker 79]. Il se base sur quatre étapes : 1) génération d'un ensemble de mots sémantiquement proches, 2) vérification de la validité de la sémantique des mots activés (stimulus) dans cet ensemble, 3) génération d'un ensemble sensoriel (aspect visuel) et 4) vérification des indices visuels dans cet ensemble. Il s'agit d'approcher à la fois physiquement et sémantiquement les mots.
- Le **modèle d'activation interactive** basé sur un réseau de neurones pour propager les indices visuels à travers les couches successives, relatives à la composition du mot : primitives, lettres et mot [McClelland 81]. La présence d'un indice visuel propage la stimulation du neurone correspondant à travers les couches des lettres et des mots. En rétro-propagation, des interactions entre les neurones activés et l'image d'entrée, sont effectuées pour aider à la décision finale.
- Le **modèle double voie** pour la reconnaissance des mots ou des pseudo-mots. La première voie procède par propagation des indices visuels pouvant mener à l'activation de mots ou de pseudo-mots. La seconde voie valide la reconnaissance des mots par une approche phonologique. Les indices visuels utilisés sont identiques à ceux utilisés dans le modèle d'activation interactive de McClelland [Côté 97] et [Pinales 01].

2.2. Systèmes perceptifs de reconnaissance

Différents systèmes perceptifs ont été repris pour la reconnaissance de mots manuscrits latins. Ces systèmes sont basés soit sur le modèle de vérification [Higgend 91], [Lecolinet 94], soit sur le modèle d'activation interactive [Pinales 01], soit sur une combinaison des deux modèles précédents [Côté 97].

- Le système *Nuscript* de Higgend associe à chaque mot du lexique un score qui est ajusté au moment de l'étape de vérification, et qui est utilisé aussi pour déduire de nouvelles caractéristiques. Il utilise les ascendants et les descendants comme primitives de base. A l'étape de vérification, il cherche à déduire d'autres primitives telles que la longueur du mot et le nombre de points correspondant à i ou j [Higgend 91].
- Le système de Lecolinet est également basé sur le modèle de vérification. Les mots sont décrits sous forme d'une liste de lettres, elles-mêmes décrites sous forme d'une liste de primitives. Les primitives utilisées sont des traits, des boucles fermées et des boucles ouvertes. Après une étape d'activation des mots candidats, le système cherche à mettre en correspondance les descriptions associées à ces mots avec l'image du mot à reconnaître. Cette mise en correspondance est effectuée de l'extérieur (caractères extrêmes) vers l'intérieur [Lecolinet 94].
- Le modèle PERCEPTO de Côté est basé sur le modèle d'activation interactive et le modèle de vérification. Il est composé de trois couches. La couche des primitives, la couche des lettres et la couche des mots. Les primitives proposées sont de deux types : les primitives primaires telles que les ascendants, les descendants et les boucles, et les primitives secondaires telles que les différentes formes et positions des boucles, la présence de la barre du t, les creux et les bosses. Les primitives primaires sont utilisées pour initialiser le système et sont propagées afin de générer un ensemble initial de mots candidats. En rétro-propagation, la présence des primitives secondaires est vérifiée suite à une mise en correspondance des mots candidats avec l'image initiale. Cette mise en correspondance est assurée par une fonction floue qui estime la position des primitives secondaires à vérifier et qui dépend de la longueur de chaque mot en nombre de lettres [Côté 97].
- Le modèle perceptif de Pinales est implémenté à l'aide d'un réseau de neurones à trois couches, basé sur le modèle d'activation interactive. Une première couche synthétise les caractéristiques extraites pour reconnaître des lettres (réseau RBF); la seconde couche prend

en entrée ces lettres et transmet ses activations vers une couche d'unités mots. Chaque caractère cursif est représenté soit par des segments de droite extraits par la méthode de Hough et des courbes ayant des orientations relativement données, soit par des points caractéristiques de l'image (extrémités, extremums de courbure, orientation et position relative du tracé), soit des ascendants, des descendants, des boucles et des concavités. Le processus de reconnaissance s'effectue de l'extérieur vers l'intérieur. Il prend en compte l'effet de régularité de chaînes de lettres et l'incertitude dans leurs positions. Pour la reconnaissance des mots, des informations additionnelles sont ajoutées telles que le nombre moyen de transitions horizontales et de boucles fermées [Pinales 01].

Le Tableau 1 présente une synthèse de ces systèmes.

Tableau 1. Systèmes perceptifs de reconnaissance de mots manuscrits latins

Systèmes	Modèles perceptifs	Primitives principales	Primitives additionnelles	spécificité
[Higgins 91]	Vérification	Ascendants et descendants	Longueur du mot et nombre de points du i	La vérification est utilisée pour déduire les primitives additionnelles
[Lecolinet 94]	Vérification	Traits, boucles fermées et boucles ouvertes		Mise en correspondance de l'extérieur vers l'intérieur
[Côté 97]	Activation interactive et vérification	Ascendants, descendants et boucles	Bosses, vallées et la barre du t	Mise en correspondance floue et vérification de la présence des primitives additionnelles
[Pinales 01]	Activation interactive	Segments de droites, courbures, extrémités et orientation du tracé, ascendant et descendant extraits par la méthode de Hough,	Nombre moyen de transition et boucles fermées	Mise en correspondance de l'extérieur vers l'intérieur, reconnaissance des lettres par un RBF

3. Description du système perceptif RNT-DF

Nous avons étendu le modèle de Côté pour l'adapter à la reconnaissance de mots arabes manuscrits, et suppléer à certaines de ses limites. Ceci s'est traduit principalement par la prise en compte des particularités de l'arabe, aux niveaux linguistique et structurel, ainsi que par la vérification au niveau local des zones des lettres. Ce réseau de neurones fait partie de la catégorie des réseaux de neurones transparents. La notion de transparence vient du fait que chaque neurone contient un concept interprétable. Les caractéristiques structurelles utilisées sont les ascendants ou « Hampes », les descendants ou « Jambages », les boucles et les points

diacritiques haut et bas. Une caractéristique de position des lettres (Isolée, au Début, au Milieu, à la Fin d'une partie du mot) est ajoutée pour tenir compte d'une des spécificités de l'écriture arabe. En effet, un mot arabe peut être composé par des parties non connexes appelées PAW (Pieces of Arabic Word). La position relative permet de rendre les caractéristiques structurelles invariantes à la largeur des mots. Dans le cas où les caractéristiques ne sont pas détectées à cause de la variabilité du manuscrit ou de leur absence dans certaines lettres arabes, deux solutions sont proposées: pendant le processus d'activation ou de propagation et en absence de toutes caractéristiques structurelles, une caractéristique « Rien » permet d'assurer l'initialisation du système en activant tout type de lettre; pendant le processus de vérification en rétro-propagation, des caractéristiques statistiques basées sur les descripteurs de Fourier, permettent d'observer localement d'une manière beaucoup plus fine les zones ambiguës sans caractéristiques structurelles. Afin de réduire la variabilité du manuscrit, une étape de normalisation locale est effectuée sur les descripteurs de Fourier [Snoussi 00].

3.1. Spécificités du RNT-DF

Comparativement aux systèmes de reconnaissance perceptive proposés, nous retraçons dans cette section les points forts de notre approche :

- Contrairement aux systèmes de [Côté 97] et [Pinales 01], le RNT-DF ne nécessite ni une phase d'apprentissage ni une large base d'exemples. En effet, les études expérimentales ont montré que l'humain ne mémorise pas toutes les déformations possibles concernant la forme d'un objet mais plutôt sa forme générale standard. Cette observation nous a conduits à faire une initialisation préalable du réseau, en associant, manuellement à chaque neurone, une description standard du concept.
- Le RNT-DF comporte une quatrième couche, entre les lettres et les mots, représentant les parties de mots connexes (PAW);
- Les caractéristiques primaires sont renforcées par la position de la lettre dans le mot (Début, Milieu, Fin ou Isolée). L'ajout de la caractéristique "RIEN" permet de résoudre le problème

des mots sans caractéristiques flagrantes (primaires). Notre système pourra ainsi traiter toutes les catégories de mots, contrairement à celui de Côté qui ne peut pas traiter les mots n'ayant pas de caractéristiques primaires.

- Une normalisation basée sur les Descripteurs de Fourier (DF) est utilisée au niveau local pour l'identification des lettres. Ces descripteurs sont normalisés et comparés à ceux de lettres imprimées, correspondant aux lettres manquantes. En effet, les Descripteurs de Fourier normalisés sont invariants par rapport à la position, à la taille, à l'inclinaison et à la rotation. Ceci permet de résoudre le problème de vérification des caractéristiques secondaires qui est une phase coûteuse et pas trop évidente dans le système PERCEPTRO.
- Une dernière contribution concerne la recherche des zones de lettres ambiguës pour la mise en correspondance entre une zone de l'image manuscrite et le mot de référence dans le lexique. Ceci pose un problème lié à deux facteurs :
 - La variabilité de la taille du mot manuscrit : nous proposons de résoudre ce problème par une normalisation globale du mot, par Fourier.
 - La difficulté de localisation précise des zones ambiguës. Nous remédions à ce problème en limitant les propositions de lettres dans la phase de rétro-propagation : soit la première ou la dernière lettre si elle n'est pas encore connue (plus simple à localiser), soit une lettre délimitée par deux zones de lettres déjà connues (aspect contextuel).

3.2. Architecture du système RNT-DF

Le système est composé de quatre couches de cellules organisées de manière hiérarchique. Une couche pour les neurones-primitives, une couche pour les neurones-lettres, une couche pour les neurones-PAWs et une couche pour les neurones-mots. La **Figure 1** décrit cette hiérarchie.

Les connexions entre les neurones du réseau se font uniquement entre les niveaux adjacents de couches. Ainsi, un neurone-mot peut être connecté à un neurone-PAW mais pas à un neurone-lettre. Les connexions entre les couches sont établies suivant des connaissances a priori. Ceci remplace la phase d'apprentissage. En effet, le RNT-DF est initialisé à partir de quatre fichiers

décrivant les primitives utilisées, la description de chaque lettre par les primitives correspondantes de chaque PAW par les lettres correspondantes et de chaque mot par les PAWs correspondants.

Chaque nœud contient une fonction d'activation, dépendant de son activation précédente et des activations pondérées en entrée du nœud, qui a la spécificité de traduire l'importance de la contribution des nœuds de la couche inférieure dans la reconnaissance de l'élément courant. Le réseau fonctionne à l'aide de deux mouvements, correspondant l'un à la propagation des activations (de la couche des caractéristiques à la couche des mots), et l'autre à la rétro-propagation contextuelle (de la couche des mots à la couche des lettres).

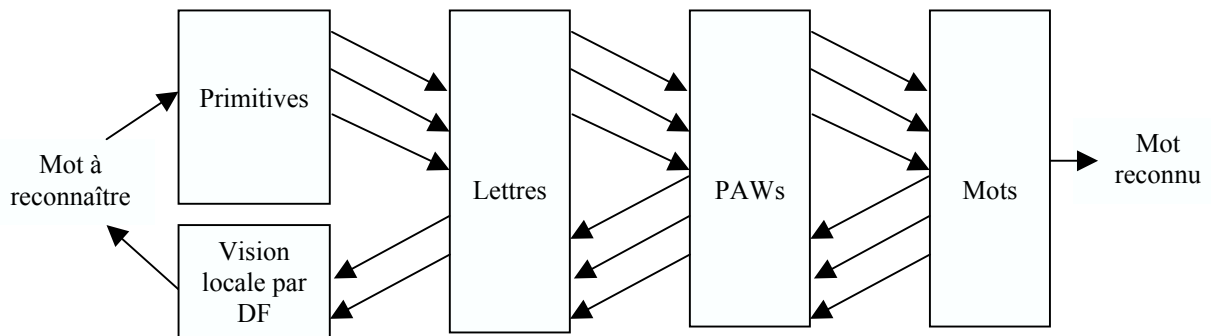


Figure 1. Architecture du RNT-DF pour la Reconnaissance de l'écriture arabe

La propagation fournit en sortie une liste de mots candidats qui seront classés selon l'ordre décroissant de leurs activations afin de décider du résultat de la reconnaissance. Une correspondance entre les zones de lettres dans l'image du mot et les lettres dans chaque mot imprimé du lexique est effectuée à la fin de chaque rétro-propagation, en se basant sur des informations contextuelles [Snoussi 02].

3.2.1. Poids de connexion

Les poids de connexion ne nécessitent pas d'étape d'apprentissage. Ils sont estimés en fonction de l'information représentée dans les neurones, par rapport à la couche inférieure. Nous faisons appel à cinq poids de connexion définis comme suit :

- En propagation :

- **Poids de connexion entre caractéristiques et lettres:** Le poids de connexion entre une caractéristique "c" et une lettre "l" est défini par la formule suivante : $\alpha_{cl} = I/NC$, où NC est le nombre de primitives extraites de l'image du mot à partir de la même zone correspondant à la lettre "l". Nous avons gardé ici la même définition que Côté. En effet, cela permet de faire varier la contribution de ces caractéristiques dans l'activation des lettres concernées.
- **Poids de connexion entre lettres et PAWs :** Le poids de connexion entre une lettre "l" et un PAW "P" est défini comme suit : $\alpha_{LP} = I/NL$, avec NL le nombre de lettres dans le PAW "P". Ici, Côté introduit la fonction floue pour estimer la position des lettres dans le mot. L'inconvénient de cette fonction est qu'elle se base sur le fait que toutes les lettres d'un mot ont la même largeur, ce qui n'est pas vrai dans le cas de l'arabe. Le problème de la position est résolu dans notre cas par les caractéristiques structurelles de la position de la lettre dans le mot (Début, Milieu, Fin ou Isolé). Donc, le poids tel que nous le définissons, présente les avantages suivants :
 - Il ne nécessite pas de phase d'apprentissage. Il peut être défini dès le début du traitement puisqu'il ne dépend que du nombre de lettres dans les PAWs, donc du lexique initial.
 - Il permet de répartir le degré de contribution à l'activation du PAW équitablement sur toutes les lettres voisines.
- En rétro-propagation :
 - **Poids de connexion entre PAWs et mots :** $\alpha_{PM} = I/NP$ est le poids entre le PAW "P" et le mot "M". NP est le nombre de PAWs dans le mot. Ce poids est établi suivant le même raisonnement que précédemment : tous les PAWs composant le mot doivent avoir la même contribution à l'activation du mot qu'ils composent.

- **Poids de connexion entre mots et PAWs:** $\alpha_{MP} = 1/NM$, est le poids entre le PAW "P" et le mot "M". NM est le nombre de mots contenant le PAW "P". Ces mots ont été activés par ce PAW et doivent contribuer à sa validation.
- **Poids de connexion entre PAWs et lettres:** $\alpha_{PL}=1/NP$, est le poids entre le PAW "P" et la lettre "L", NP est le nombre de PAWs activés par la lettre "L".

3.2.2. Degrés d'activation

Ils se mettent à jour durant le traitement en tenant compte d'une équation définie dans [McClelland 81]. Chaque neurone a un degré d'activation $A_i(t)$ qui varie entre 0 et 1 suivant

$$\text{l'équation (1) : } A_i(t+\Delta t) = (1 - \theta) A_i(t) + \sum_j \alpha_{ij} a_j(t) (M - A_i(t)) \quad (1)$$

Avec $A_i(t+\Delta t)$: la nouvelle valeur d'activation de la cellule au temps $(t+\Delta t)$, $\theta = 0.07$ est une constante de décroissance de l'activation, $M = 1$ est le degré d'activation maximale, $a_j(t)$ est l'activation d'un voisin actif (excité), et α_{ij} est le poids de la connexion entre les cellules i et j .

Le facteur $(1 - A_i(t))$ ajuste la contribution du voisin j afin de garder le degré d'activation au dessus d'un certain seuil (maximum). En effet, lorsque l'activation atteint sa valeur maximale=1; l'effet des voisins est réduit à zéro. Ainsi, l'activation est toujours bornée [Côté 97].

4. Reconnaissance locale des lettres

Nous faisons appel à la normalisation locale dans le cas où certains caractères ne présentent pas de caractéristiques primaires ou bien si ces caractéristiques n'ont pas pu être extraites à cause de la variabilité de l'écriture manuscrite. Le problème rencontré dans ce traitement local est la délimitation des zones de normalisation.

4.1. Délimitation des zones ambiguës

La délimitation des zones ambiguës traite les PAWs séparément. Elle est effectuée en se basant sur les minimums globaux du contour supérieur et en tenant compte des zones déjà extraites par l'étape d'extraction des caractéristiques structurelles visuelles. Les zones prises en compte lors de la normalisation sont ceux au début, à la fin, isolés ou entre deux zones déjà extraites. La

Figure 2 présente des exemples de délimitation de zones ambiguës, à la fin du deuxième PAW (a), au début du premier PAWs (b) et entre deux caractéristiques : Point diacritiques haut au début et boucle au milieu d'un mot composé par un seul PAW (c) .

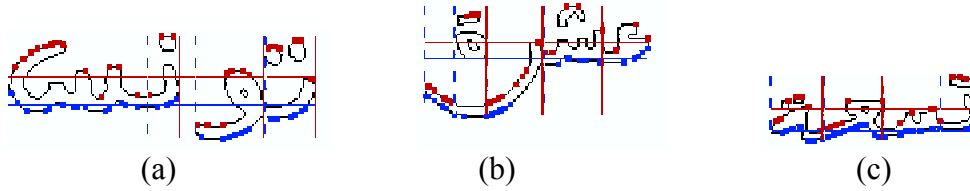


Figure 2. Résultats expérimentaux de délimitation de zones à normaliser

4.2. Normalisation des descripteurs de Fourier locaux

Le choix des descripteurs de Fourier est dû à leurs propriétés d'invariabilité, de stabilité et de complétude suite à une étape simple de normalisation. En effet, à partir de la fonction du contour d'une image, on peut générer un ensemble complet de nombres complexes dits Descripteurs de Fourier ou harmoniques. Ces descripteurs représentent la forme dans le domaine fréquentiel.

- Les descripteurs de basse fréquence décrivent la forme générale.
- Les descripteurs de haute fréquence décrivent les détails.

La **Figure 3** présente une illustration graphique des descripteurs de basse fréquence, premier et second harmonique, décrivant l'orientation, la taille et la position de la forme et les descripteurs de hautes fréquences, décrivant les détails de la lettre « ssin ».

Pour la classification, un sous-ensemble de descripteurs de Fourier est souvent suffisant pour discriminer les différentes formes. Puisque les applications pratiques utilisent les données discrètes, il est raisonnable d'employer la Transformation de Fourier discrète. L'équation (2) décrit la transformée de Fourier de la fonction du contour de longueur N , avec A_0 et C_0 les coefficients de Fourier d'ordre 0 et a_n, b_n, c_n, d_n les descripteurs de Fourier d'ordre n .

$$\begin{aligned}
 X_N(k) &= A_0 + \sum_{n=1}^N a_n \cos \frac{2n\pi k}{N} + b_n \sin \frac{2n\pi k}{N} \\
 Y_N(k) &= C_0 + \sum_{n=1}^N c_n \cos \frac{2n\pi k}{N} + d_n \sin \frac{2n\pi k}{N}
 \end{aligned} \tag{2}$$

L'approximation du contour par les coefficients de Fourier peut être vue comme une superposition de formes elliptiques de différentes tailles, positions et orientations.



Figure 3. L'approximation du contour de la lettre « ssin » par les 20 premiers harmoniques

Etant donné que les descripteurs de basse fréquence donne une idée sur la forme globale de la lettre, les étapes de normalisation se basent sur les descripteurs d'ordre 0 et d'ordre 1 pour la normalisation de la position, de la taille et de l'orientation. Ces étapes de normalisation sont présentées en détail dans [Snoussi 00] et [Snoussi 02].

- **Normalisation de la position :** Elle est assurée en ignorant les termes A_0 , C_0 de l'équation 0. Ceci correspond à un changement de repère par rapport au centre de gravité de la lettre décrite par ces deux termes.
- **Normalisation de la taille :** Elle consiste à diviser tous les coefficients d'une lettre par le module de l'axe principal du premier harmonique.
- **Normalisation de l'orientation :** L'angle d'orientation de la lettre est aussi déduit de l'axe principal du premier harmonique décrit par la formule (3).

$$\psi_1 = \arctan \left[\frac{y_1^*(0)}{x_1^*(0)} \right] = \arctan \frac{c_1^*}{a_1^*} \quad (3)$$

4.3. Décision sur la lettre insérée

L'objectif de la normalisation locale est de créer une relation de ressemblance entre la zone ambiguë et les lettres imprimées proposées par la phase de rétro-propagation du RNT. Il a pour but d'aider le RNT-DF à décider de l'un des caractères proposés afin qu'il soit inséré dans le nouveau processus de propagation. Suite à la normalisation des coefficients de Fourier aussi bien de la zone ambiguë que des lettres proposées, un calcul de distance est effectué entre cette zone et chacune des lettres de référence. Cette distance est définie par l'équation (4) avec "ma", la lettre manuscrite et "Im", la lettre imprimée [Snoussi 01], [Snoussi 02].

$$D(ma, Im) = \frac{\sum_{i=1}^N (a_{ma}^2 - a_{Im}^2) + (b_{ma}^2 - b_{Im}^2) + (c_{ma}^2 - c_{Im}^2) + (d_{ma}^2 - d_{Im}^2)}{\sum_{i=1}^N (a_{ma}^2 + a_{Im}^2 + b_{ma}^2 + b_{Im}^2 + c_{ma}^2 + c_{Im}^2 + d_{ma}^2 + d_{Im}^2)} \quad (4)$$

5. Expérimentation

Une évaluation expérimentale a concerné aussi bien l'étape de normalisation locale que l'étape de reconnaissance globale par le RNT. L'expérimentation a été appliquée sur deux types de lexiques. Un lexique de 71 mots de montants littéraux et un lexique de 150 mots de noms de villes tunisiennes. Le lexique des montants comporte 4970 mots pris au hasard d'une base de données de 5000 chèques omni-scripteurs, collecté au sein du labo LSTS. Le lexique des noms de villes tunisiennes est pris de la base de données IFN/ENIT, comportant 22000 mots omni-scripteurs collectés manuellement (dans le cadre d'un projet de collaboration entre les écoles IFN en Allemagne et l'ENIT en Tunisie). Pour chaque mot des deux lexiques, nous disposons de 70 échantillons.

Dans le **Tableau 2**, nous présentons quelques résultats de la décision locale après normalisation des descripteurs de Fourier. Nous remarquons que la distance calculée entre les différentes classes est toujours minimale (diagonale) pour la lettre adéquate. Cette évaluation a été effectuée sur toutes les lettres arabes pré-segmentées.

Tableau 2. Résultats expérimentaux de la normalisation locale

Im						Im					
ma	0,7398	0,7815	0,9287	0,9644	1,0000	ma	0,0953	0,2552	0,9123	0,8888	0,9847
na	0,0366	0,0239	0,0398	0,7103	0,8073	na	0,3413	0,0726	0,9034	0,8480	0,8203
ca	0,1377	0,1030	0,0233	0,6893	0,6496	ca	0,9263	0,9168	0,0033	0,0255	0,1246
sa	0,6708	0,6415	0,6766	0,3140	0,4520	sa	0,8799	0,8217	0,0659	0,0210	0,1530
sa	0,7664	0,7792	0,7926	0,4909	0,1452	sa	1,0000	0,8650	0,0981	0,0904	0,0575

Dans le Tableau 3, nous présentons les taux de reconnaissance obtenus pour les différents lexiques, et ceci avant et après intégration de la normalisation locale. Nous remarquons que le taux augmente après intégration de la normalisation. Le taux de reconnaissance des noms de

viles est effectu   uniquement pour le syst  me RNT-DF entier. Nous remarquons que le taux de reconnaissance d  cro  t car la taille du lexique a augment  , donc la probabilit   de confusions entre mots a aussi augment  .

Tableau 3 : Taux de reconnaissance du RNT

Syst��mes	Taille du Lexique	Taille de la base	Nombre de scripteurs	Taux de reconnaissance
RNT	71	4970	98	90%
RNT-DF	71	4970	98	97%
RNT-DF	150	10500	445	95%

6. Conclusion et perspectives

Nous avons pr  sent   dans cet article un mod  le perceptif de reconnaissance de mots arabes manuscrits bas   sur un R  seaux de Neurones    repr  sentation Transparente des donn  es du lexique (RNT). Il op  re par une premi  re vision globale des caract  ristiques structurelles apparentes, faciles    d  tecter. En cas de difficult  s d'extrac  tion de certaines de ces caract  ristiques, une vision locale par des Descripteurs plus fins de Fourier (DF) permet d'aider le RNT-DF    une meilleure d  cision. Nous avons choisi d'utiliser les descripteurs de Fourier normalis  s car ils sont invariants    tout type de distorsion de taille, d'orientation et de position.

L'information contextuelle, apport  e par l'  tape de r  tro-propagation, peut   tre exploit  e pour am  liorer l'  tape d'extrac  tion des caract  ristiques structurelles comme par exemple l'extrac  tion de la ligne de base. Vu l'invariabilit   de ces descripteurs, nous pensons dans une deuxi  me perspective, utiliser la vision locale au niveau de PAWs afin d'  viter l'  tape de d  limitation des zones. Nous pr  voyons   galement d'  tudier l'apport de l'apprentissage, tout comme les syst  mes de Ct   et Pirales, pour le RNT-DF. Une pond  ration des entr  es du RNT sera   galement effectu  e afin de privil  gier l'influence de certaines caract  ristiques.

Une   tude comparative du RNT-DF avec les syst  mes perceptifs existant ainsi que les syst  mes de reconnaissance de mots manuscrits arabe est enfin envisag  e..

7. Bibliographie

- [Bekker 79] C.A. Becker, Semantic Context and Word Frequency Effects in Visual Word Recognition, *Journal of Experimental Psychology : Human Perception and Performance*, vol 5 pp 252-259, 1979.
- [Côté 97] M.Côté, Utilisation d'un Modèle d'Accès Lexical et de Concepts Perceptifs pour la Reconnaissance d'Images de Mots Cursifs, Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications de Paris, 1997.
- [Higgins 91] C.A. Higgins and P.E. Bramall, Handwriting Recognition based on the human reading process, *ICDAR'91*, vol 1, N°1, pp 115-167, 1991.
- [Lecolinet 94] E. Lecolinet, Cursive Script Recognition by backward matching, *Advances in Handwriting and Drawing : A Multidisciplinary Approach*, pp 117-135, 1994.
- [McClelland 81] J.L. McClelland and D.E. Rumelhart, An interactive Activation Model of Contextual Effects in Letter Perception, *Psychological Review*, pp 375-407, 1981.
- [McClelland 85] J.L. McClelland and D.E. Rumelhart, Distributed Memory and the Representation of General and Specific Information, *Journal of Experimental Psychology : General*, pp 159-188, 1985.
- [Pinales 01] J.R.Pinales, Reconnaissance hors-ligne de l'écriture cursive par l'utilisation de modèles perceptifs et neuronaux, Thèse de Doctorat, Ecole Nationale Supérieure des Télécommunications de Paris, 2001.
- [Snoussi 00] S.Snoussi.Maddouri, H.Amiri and A.Belaid, Local Normalization Toward Global Recognition, *Document Analysis System IAPR-DAS'00*, Brezil, 2000.
- [Snoussi 02] S.Snoussi.Maddouri, H.Amiri, A.Belaid and C.Choisy, Combination of Local and Global Vision Modeling for Arabic Handwritten Word Recognition, *International Workshop Frontier in HandWriting IWFHR'02*, Canada, 2002.