UNIVERSITE KASDI MERBAH OUARGLA

Faculté des nouvelles technologies de l'information et de la communication

Département d'électronique et communications

Mémoire

MASTER PROFESSIONNEL

Domaine : Sciences et technologies
Filière : Electronique
Spécialité : Instrumentation

Présenté par :

MEGAR Ishak et GEUDDA Adel

Thème

Contribution à la reconnaissance d'écriture arabe manuscrite en utilisant la classification neuronale

Soutenu publiquement
le : 02/06/2016

Devant le jury :

M. MERAOUMIA Abdellah	MC(B)	Président	UKM Ouargla
M. BENCHABANE Abderrazak	MC(B)	Encadreur	UKM Ouargla
Mme BENKRINAH Sabra	MA(A)	Examinateur	UKM Ouargla
M. SAIGAA Mohamed	MA(A)	Examinateur	UKM Ouargla

Année Universitaire : 2015 /2016

Remerciements

Avant toute chose, nous remercions dieu le tout puissant de nous avoir aidé et éclairé le chemin pour la réalisation de ce mémoire.
Nous voudrons remercier Dr BENCHABANE Abderrazak pour nous avoir encadré.

Nous tenons à remercier le Dr. A. MERAOUMIA, qui nous a fait l'honneur de présider le jury de notre mémoire. Nous remercions également les membres de jury, Mme S. BENKRINAH et Mr M. SAIGAA pour le temps qu'ils ont bien voulu consacrer à lire et à juger le travail en tant que examinateurs de notre mémoire.

Nous adressons nos remerciements à tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce mémoire.
Enfin, nous voudrons remercier notre familles qui nous ont donné l'éducation sans laquelle nous ne serons pas là aujourd'hui.

Liste des figures

1.1 Types d'écritures…….…….……...…..... 6
1.1 Difficultés de l'écriture…….. 7
1.3 Principe de la reconnaissance de caractères ou de mots isolés …… 8
1.4 Système de reconnaissance de caractères ou de mots isolés…… 9
1.5 Inclinaison globale du mot par rapport au cadre de l'image..... 10
1.6 Inclinaison des lettres par rapport à l'axe moyen du mot.…… 10
1.7 Exemple du mot avant et après squelettisation. ……. 11
1.8 Résultats de la normalisation ……..…… 11
1.8 Exemple de localisation des lignes d'écriture……. 11
1.9 Histogramme de la segmentation de mots…….. 12
2.1 Exemple de projection……. 15
2.2 Différentes zones……. 15
2.3 Différentes cavités ……..... 16
2.4 Modèle de neurone formel …….… 21
2.5 Perceptron …… 21
2.6 Perceptron multicouche fortement connecté aux couches cachées……. 22
3.1 Exemple de chiffres normalisés extraits de MNIST ……..……. 27
3.2 Système de reconnaissance utilisé…… 29
3.3 Exemples d'images extraites de la base AHDB…… 31
3.4 Caractéristiques utilisées…… 32

Liste des tableaux

2.1. Moments de Hu du chiffre ‘5’ avec differentes transformations……. 18

2.2 Moments de zernike du chiffre ‘5’ avec differentes transformations…….. 20

3.1 Taux d’apprentissage et de reconnaissance pour la base MNIST…… 29

3.2 Taux d’apprentissage et de reconnaissance avec combinaisons de caractéristiques….. 30

3.3 Matrice de confusion de l’apprentissage ……... 30

3.4 Matrice de confusion de la reconnaissance……. 31

3.5 Taux d’apprentissage et de reconnaissance pour les mots isolés…… 32

Table des matières

Remerciements…….…..... i

Liste des figures…….…… ii

Liste des tableaux…….…….… iii

Table des matières…….……..…….. Iv

Introduction Générale…….…….. 1

Chapitre 1

Généralités sur la reconnaissance de l'écriture manuscrite

1.1. Introduction 3
1.2. Types de reconnaissances 3
1.2.1. Reconnaissance de caractères 3
1.2.2. Reconnaissance de mots 4
1.3. Reconnaissance de documents 4
1.4. Ecriture arabe 5
1.5. Classes des approches de reconnaissance de l'écriture 5
1.6. Complexité d'un système de reconnaissance 6
1.6.1. Disposition spatiale du texte 6
1.6.2. Nombre de scripteurs 6
1.6.3. Taille du vocabulaire 7
1.7 Applications de la reconnaissance de l'écriture……7
1.8. Architecture d'un système de reconnaissance de caractères et de mots 8
1.8.1. Mode d'acquisition en ligne/ hors ligne 8
1.8.2. Prétraitements 8
1.8.3. Extraction des caractéristiques 11
1.8.4. Classification 12
1.8. Conclusion 12

Chapitre 2

Extraction des caractéristiques et classification

2.1. Introduction 13
2.2. Extraction des caractéristiques 13
2.2.1. Approche VH2D (vertical-horisontal-2diagonal) 13
2.2.2. Zonage 14
2.2.3. Profils 15
2.2.4. Cavités 15

2.2.5. Moments géométriques 15
2.2.6. Moments de Zernike 18
2.3. Classification 19
2.3.1. Neurone formel 19
2.3.2. Perceptron 20
2.3.3. Perceptron multicouches 20
2.3.4. Règles d'apprentissage 21
2.3.5. Rétro-propagation du gradient 21
2.3.6. Convergence et adaptation des poids 22
2.3.7. Algorithme de rétro-propagation 23
2.3.8. Critère d'arrêt 23
2.4. Application des réseaux de neurones à la reconnaissance de l'écriture 23
2.5. Conclusion 24

Chapitre 3

Application du perceptron multicouche à la reconnaissance de caractères arabes

3.1. Introduction……25
3.2. Reconnaissance des chiffres arabes……25
3.2.1. Système de reconnaissance proposé……25
3.2.2. Résultats de simulation……28
3.3. Reconnaissance de mots arabes isolés à vocabulaire limitée……30
3.3.1. Caractéristiques utilisées……30
3.3.2. Résultats de simulation……31
3.4. Conclusion…… 32
Conclusion générale……33
Bibliographie……34

Introduction générale

La reconnaissance de l'écriture relève du domaine de la reconnaissance des formes qui s'intéresse aux formes de caractères, 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 manuscrite est aujourd'hui en plein d'essors et les domaines d'applications ne cessent de se multiplier (l'enregistrement des chèques bancaires, le traitement automatique de dossiers administratifs) [5]. La reconnaissance de l'écriture manuscrite s'avère un problème extrêmement complexe qui n'a pas de solution satisfaisante à ce jour. Ceci à cause des difficultés auxquelles sont confrontés les chercheurs et les développeurs, telles que la variabilité de la forme, du style, et de l'inclinaison de l'écriture. L'écriture manuscrite arabe est naturellement cursive, difficile à traiter, et présente une grande variabilité [12].

Dans le domaine de la reconnaissance de l'écriture, les caractéristiques peuvent être décrites comme un moyen permettant de distinguer un objet (mot, lettre, chiffre) d'une classe d'un autre objet (mot, lettre, chiffre) d'une autre classe [4].

Les primitives sont généralement définies par expérience ou par intuition. Plusieurs caractéristiques peuvent être extraites. La représentation des caractéristiques utilisée est une représentation vectorielle. La taille du vecteur peut être large si un grand nombre de caractéristiques est extrait. Il a été observé que certaines caractéristiques extraites sont non pertinentes ou redondantes au système de reconnaissance. Donc, il est nécessaire de définir des caractéristiques discriminantes lors du développement d'un système de reconnaissance.

Les réseaux de neurones sont massivement parallèles et ont la capacité de mémoriser la connaissance à partir d'exemples qui proposent une solution intéressante aux problèmes de classification [1]. Le concepteur constitue une base d'apprentissage décrivant diverses situations auxquelles seront confrontés les systèmes. Au terme d'un processus d'apprentissage utilisant uniquement ces exemples, le réseau se révèle capable de traiter un grand nombre de situations relatives au problème considéré. Cette propriété, appelée reconnaissance, explique l'engouement pour de tels systèmes. Nous souhaitons, dans ce mémoire, réaliser le module de classification à l'aide d'un système neuronale. L'apprentissage et la reconnaissance s'effectueront évidement à partir des informations recueillies par l'algorithme d'extraction des caractéristiques.

L'objectif de notre travail consiste à développer un système de reconnaissance des caractères manuscrits arabes utilisant les réseaux de neurones.

Notre mémoire se subdivise donc comme suit:

Le premier chapitre est consacré à l'état de l'art dans le domaine de la reconnaissance de l'écriture manuscrite où nous décrivons les étapes nécessaires au développement d'un système de reconnaissance. Le deuxième chapitre se répartie en deux sections, la première étudie brièvement l'étape de l'extraction des caractéristiques, et la seconde présente l'étape de classification par les réseaux de neurones où nous décrivons la théorie des réseaux de neurones. A la fin du chapitre, l'application de ces modèles pour la reconnaissance de l'écriture manuscrite est expliquée. Le troisième chapitre décrit les étapes de la reconnaissance appliquées dans notre système ainsi que les résultats obtenus.

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

Chapitre 1

Généralités sur la reconnaissance de l'écriture manuscrite

1.1. Introduction

La reconnaissance automatique de l'écriture consiste à la création des systèmes capables de reconnaitre des caractères manuscrits ou imprimés. Son principe est de transformer un texte écrit en une représentation compréhensible et reproductible par la machine. Malgré le grand progrès de la reconnaissance automatique de l'écriture, ce domaine demeure très actif vu à la grande variabilité de l'écriture manuscrite [6].

1.2. Types de reconnaissances

1.2.1. Reconnaissance de caractères

La reconnaissance de caractères concerne la reconnaissance de caractères numériques ou alphabétiques. Les caractères sont donnés par des images étiquetées, par exemple une image du caractère '5' aura une étiquette 5.

C'est la tâche la plus basique d'un système de reconnaissance de l'écriture. L'effort d'analyse est concentré sur un seul élément à la fois du vocabulaire (vue comme une forme globale).

Ces études portent sur la reconnaissance de caractères manuscrits provenant de la segmentation de chiffres comme par exemple le montant numérique d'un chèque ou le code postal d'une adresse ou de la segmentation en lettres minuscules et/ou majuscules dans le cas par exemple du montant littéral d'un chèque ou du nom de la ville figurant sur une adresse. Ces caractères isolés présentent de fortes variations principalement provoquées par la position de la lettre dans le mot.

Deux types de reconnaissance de caractères numériques manuscrits : la reconnaissance des chiffres isolés : peut être considéré comme un problème de classification des images des caractères isolés dans un ensemble d'alphabet donné. La reconnaissance des chaînes numériques manuscrites : il est à noter que ce type des systèmes commence généralement par une phase de segmentation qui consiste à séparer la chaîne numérique à des entités (chiffres) isolées et reconnaître ces dernières par un module de reconnaissance des chiffres isolés[3].

1.2.2. Reconnaissance de mots

Dans la présente situation, on s'intéresse à la reconnaissance du mot entier. Deux types de reconnaissance peuvent être considérés, la reconnaissance globale où le processus de la reconnaissance est ramené à un processus de reconnaissance de caractères avec des formes plus complexes que les caractères. Le deuxième type est une reconnaissance analytique où le mot est divisé en graphèmes. Elle cherche à identifier les caractères ou les sous-caractères (graphèmes) issus de la segmentation (séparation de mots, des caractères) pour reconstituer les mots.

La difficulté d'une telle approche a été clairement évoquée par Sayre en 1973 et peut être résumée par le dilemme suivant : "*pour reconnaître les lettres, il faut segmenter le tracé et pour segmenter le tracé, il faut reconnaître les lettres*" [3]. Cette approche est la seule applicable dans le cas de grands vocabulaires. Elle peut s'adapter facilement à un changement de vocabulaire. Elle permet théoriquement une discrimination plus fine des mots car elle se base sur la reconnaissance des lettres qui la composent et il est possible de récupérer l'orthographe du mot reconnu. Son inconvénient principal demeure la nécessité de l'étape de segmentation avec les problèmes de sous- ou de sur-segmentation que cela implique.

Certaines des approches actuelles se proposent de tirer avantage des deux méthodes, réduisant la complexité de l'approche globale en l'appliquant sur des entités plus petites (lettres). L'approche analytique recherche la séquence de lettres contenues dans l'image à reconnaître. Certains modèles permettent de combiner ces deux niveaux en un seul et peuvent ainsi s'affranchir de la segmentation préalable de l'image

1.3. Reconnaissance de documents

Cette approche a une vision générale du mot; elle se base sur une description unique de l'image du mot, vue comme une entité indivisible. Disposant de beaucoup d'informations, elle absorbe plus facilement les variations au niveau de l'écriture.

Cette approche présente l'avantage de garder le caractère dans son contexte avoisinant, ce qui permet une modélisation plus efficace des variations de l'écriture et des dégradations qu'elle peut subir. Cependant cette méthode est pénalisante par la taille mémoire, le temps de calcul et la complexité du traitement qui croient linéairement avec la taille du lexique considéré, d'où une limitation du vocabulaire souvent appliquée pour réduire la liste de mots candidats dans le contexte d'une reconnaissance à grands vocabulaires (mais limitée dynamiquement : pouvant être réduit de façon dynamique, comme l'ensemble des noms de rues associés à un bureau de poste distributeur).

1.4. Ecriture arabe

L'alphabet arabe comprend vingt-neuf lettres fondamentales (vingt-huit si l'on exclut la hamza, qui se comporte, soit comme une lettre à part entière, soit comme un diacritique). Le sens d'écriture va de droite à gauche. Il n'y a pas de différence entre les lettres manuscrites et les lettres imprimées ; les notions de lettre capitale et lettre minuscule n'existent pas. En revanche, la plupart des lettres s'attachent entre elles, même en imprimerie, et leur graphie diffère selon qu'elles sont précédées et/ou suivies d'autres lettres ou qu'elles sont isolées (variantes contextuelles). Certaines lettres, cependant, ne s'attachent jamais à la lettre suivante de ce fait, un mot unique peut être entrecoupé d'un ou plusieurs espaces, lesquels sont aussi utilisés pour séparer les mots. La longueur de cet espace inter-mot est généralement supérieure à l'espace intra-mot entre caractères non attachés [12,13].

L'écriture arabe est un cas très intéressant en reconnaissance automatique de l'écriture et du document suite à sa cursivité. De plus, les diacritiques rendent la reconnaissance encore plus délicate. Les techniques utilisées sont celles de la reconnaissance de mots du paragraphe même si le problème de la segmentation n'est pas encore totalement résolu. Enfin les systèmes actuels ne traitent que les écrits sans vocalisation donc sans diacritique.

1.5. Classes des approches de reconnaissance de l'écriture

Le problème de la reconnaissance de l'écriture peut être abordé par deux approches différentes, approche hors-ligne et approche en-ligne, ces deux approches se diffèrent par la façon d'acquérir les tracés de l'écriture (en-ligne ou hors-ligne), et par la présence ou non des informations décrivant l'ordre de construction de chaque tracé; donc les méthodes de reconnaissance se diffèrent d'une approche à une autre [3].

1.5.1. Reconnaissance hors ligne ou statique

Dans le cas hors ligne, il s'agit de reconnaître des textes manuscrits à partir de documents écrits au préalable. L'image du texte écrit est numérisée à l'aide d'un scanner, les informations recueillies se présentent sous la forme d'une image discrète constituée d'un ensemble de pixels. L'écriture prend l'aspect d'un signal spatial bidimensionnel numérisé.

1.5.2. Reconnaissance en ligne ou dynamique

Dans le cas en ligne, il s'agit de reconnaître l'écriture au fur et à mesure de son tracé. Le texte est saisi avec un stylo et une tablette à numériser, les informations recueillies sont constituées par une suite ordonnée de points (définis par leurs coordonnées) échantillonnées à cadence fixe. L'écriture prend l'aspect d'un couple de signaux temporels numérisés. La

reconnaissance en-ligne se fait en temps réel, c'est à dire, pendant l'écriture (le temps de reconnaissance dépend de la technique utilisée et de la vitesse de l'ordinateur).

La reconnaissance en-ligne est généralement beaucoup plus efficace que la reconnaissance hors-ligne car les échantillons sont beaucoup plus informatifs. En revanche, elle nécessite un matériel beaucoup plus couteux et impose de fortes contraintes au scripteur puisque la capture de l'encre doit se faire au moment de la saisie (capture synchrone) et non a posteriori (capture asynchrone) comme les Smartphone, les Tablet PC…. Ces systèmes sont utilisés dans plusieurs équipements électroniques[3].

1.6. Complexité d'un système de reconnaissance

1.6.1. Disposition spatiale du texte

La disposition spatiale du texte joue un rôle important dans les systèmes de reconnaissance. Une écriture pré-casé ou guidée par des cadres, comme les formulaires par exemple est plus aisée à la reconnaissance qu'une écriture cursive.

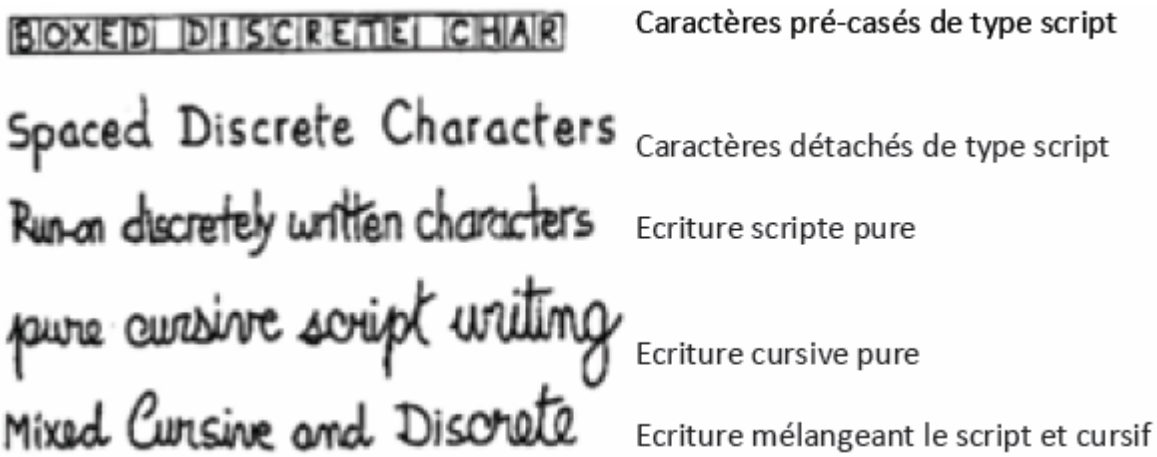

Figure 1.1 Types d'écritures [3].

1.6.2. Nombre de scripteurs

La difficulté de reconnaissance croît avec le nombre de scripteurs. On peut distinguer trois types de systèmes de reconnaissance d'écriture, classés par ordre de complexité de fonctionnement croissante :

M**ono-scripteur** : un seul scripteur peut utiliser le système de reconnaissance après apprentissage de son écriture ;

M**ulti-scripteur** : le système peut reconnaître les écritures d'un groupe restreint de personnes, soit après adaptation à l'écriture de chacun, soit sans adaptation ;

Omni-scripteur : le système est censé reconnaître toutes les écritures. Dans ce cas, la variabilité intra-scripteur s'ajoute à la variabilité inter-scripteur.

1.6.3. Taille du vocabulaire

De la même manière que le nombre des scripteurs, la difficulté de la reconnaissance croît avec la taille vocabulaire. Trois grandes catégories d'applications peuvent être distinguées :

Applications à vocabulaire très limité : où le nombre de mots (ou de symboles) à reconnaître constitue un lexique de taille réduite (inférieure à 100 mots), comme par exemple dans le cas de l'ensemble des mots utilisés pour écrire en toutes lettres les montants des chèques. Il est alors possible de confronter chaque mot à reconnaître à l'ensemble des mots du lexique ;

Applications à vocabulaire étendu : mais pouvant être réduit de façon dynamique, comme l'ensemble des noms de rues associés à un bureau de poste distributeur ;

Applications à vocabulaire très étendu : plusieurs milliers ou dizaines de milliers de mots formant un dictionnaire et pour lesquels la confrontation systématique n'est plus possible comme par exemple le dictionnaire des noms de commune.

La figure ci-dessous montre la difficulté de la reconnaissance automatique des mots ou caractères manuscrits en fonction de la disposition du texte, le nombre de scripteurs et la taille de vocabulaire.

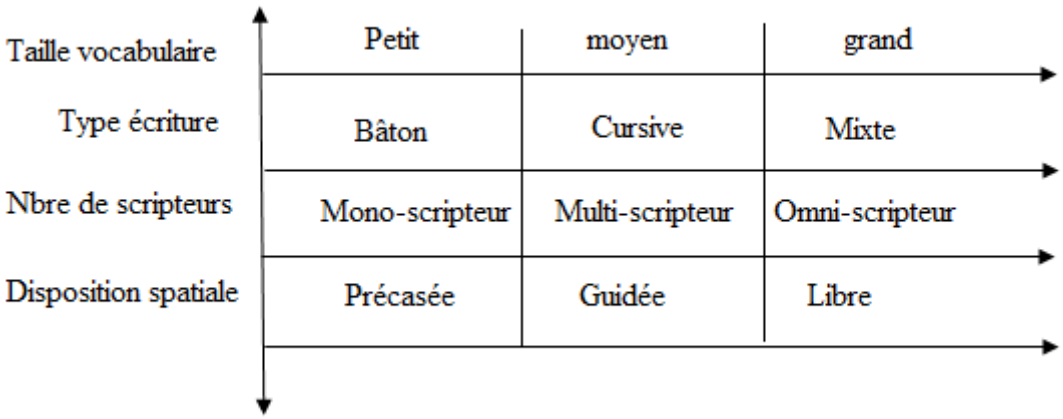

Figure 1.2. Difficultés de l'écriture [3].

1.7. Applications de la reconnaissance de l'écriture

La reconnaissance de l'écriture a permis le développement d'applications particulières et efficaces concernant la reconnaissance de l'écriture hors-ligne et en-ligne. Pour la reconnaissance de l'écriture hors ligne, elle connaît un essor important dans les domaines

associés au développement d'intérêts économiques comme la lecture des adresses postales, la reconnaissance des montants littéraux manuscrits, la reconnaissance des montants numériques et la lecture des formulaires de sondage. Quant à la reconnaissance de l'écriture en ligne, elle connaît un essor important dans les domaines associés à une communication homme-machine naturelle. Elle a pour ambition de remplacer le clavier et la souris de l'ordinateur par un stylo (ordinateurs sans clavier) [5].

1.8. Architecture d'un système de reconnaissance de caractères et de mots

Tout système de reconnaissance de caractères fait appel généralement aux étapes suivantes: l'acquisition des donnés, les prétraitements, la segmentation, l'extraction des caractéristiques, et enfin la classification [3] (Figure 1.3).

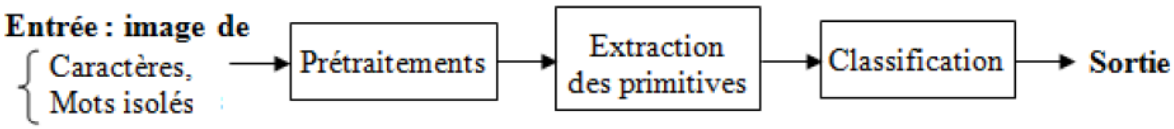

Figure 1.3 : Principe de la reconnaissance de caractères ou de mots isolés

1.8.1. Mode d'acquisition en ligne/ hors ligne

Les méthodes de reconnaissance se regroupent en deux catégories selon le mode d'acquisition de l'écriture.

L'écriture en ligne où le mot est obtenu par une saisie en continue utilisant des dispositifs adéquats. Le mot obtenu est représenté par une séquence de points décrivant la trajectoire du stylet dans le temps.

L'écriture hors-ligne (ou en différé, ou encore statique) est obtenue par la saisie d'un texte déjà existant, obtenue par un scanner ou une caméra. Dans ce cas, on dispose d'une image binaire ou en niveaux de gris. Ce mode introduit une difficulté relative à la variabilité du tracé en épaisseur et en connectivité, nécessitant l'application de techniques de prétraitement.

1.8.2. Prétraitements

La phase des prétraitements a pour but la réduction du nombre de données pour n'en garder que les informations utiles. Il s'agit de traitements dits de bas niveau (binarisation, réduction du bruit, segmentation… etc.) [6, 13, 14].

Figure 1.4 : Système de reconnaissance de caractères ou de mots isolés.

Filtrage

L'image des caractères peut être entachée de bruits dus aux artefacts de l'acquisition et à la qualité du document, conduisant soit à une absence de points ou à une surcharge de points.
La réduction du bruit consiste à détecter et à éliminer les pixels qui le représentent. Le lissage et la suppression de bruit peuvent être obtenus par un filtrage. Le filtrage est une opération de voisinage, la valeur d'un pixel est remplacée par la valeur d'une fonction appliquée à ce pixel et à ses voisins. Dans le domaine de reconnaissance de l'écriture, plusieurs types de filtrage peuvent être considérés selon le type de bruit rencontré ou le traitement désiré [12].

Binarisation

La binarisation est une opération qui permet le passage d'une image de niveaux de gris à une image binaire composée de 2 valeurs 0 et 1, plus simple à traiter.
Lorsque les documents papier sont de bonne qualité et de fond blanc, un simple seuillage global suffit après analyse de l'histogramme des niveaux de gris. Tous les pixels plus lumineux que le seuil sont pris comme blancs, les autres comme noirs. Mais lorsque le fond est texturé ou si le document est dégradé (pliures, taches ...), une analyse plus fine est nécessaire

Redressent de la ligne de base

Un mauvais positionnement du document sur le scanner, conduit à une inclinaison de l'image. Pour une bonne exploitation des documents, il faut donc détecter l'angle d'inclinaison et redresser les images, de telle sorte que les lignes de texte soient parallèles aux bords. L'angle du document (ou de la ligne), θ, est celui qui correspond a l'histogramme d'entropie maximale.
pour corriger cette inclinaison, il suffit d'appliquer une rotation de l'image d'angle θ [3, 14]:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.1}$$

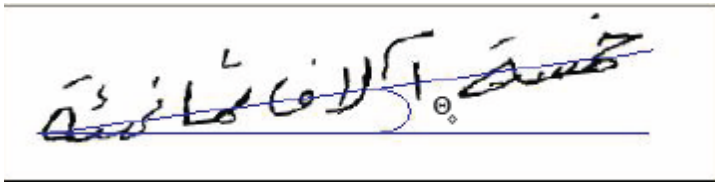

Figure 1.5. Inclinaison globale du mot par rapport au cadre de l'image[14].

Redressement des écritures penchées

Cette technique facilite la segmentation préalable des mots en caractères. L'idée est de trouver l'angle moyen d'inclinaison puis de faire une transformation géométrique de type cisaillement de l'image

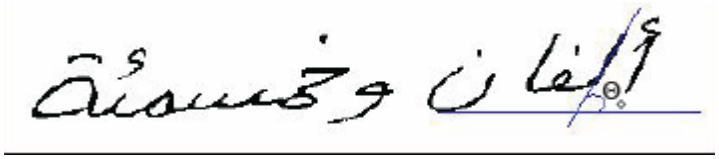

Figure 1.6. Inclinaison des lettres par rapport à l'axe moyen du mot [14].

Squelettisation

La squelettisation sert à obtenir une épaisseur égale à un pixel du trait d'écriture et de se ramener ainsi à une écriture linéaire. Le squelette doit préserver la forme, connexité, topologie et extrémités du tracé, et ne doit pas introduire d'éléments parasites ; et donc simplifier la représentation.

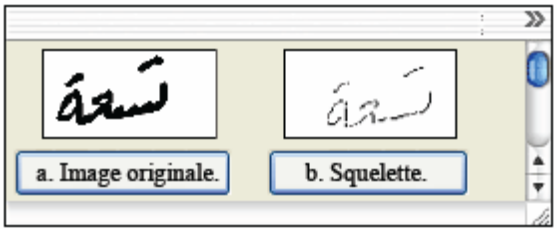

Figure 1.7. Exemple du mot avant et après squelettisation [14].

Normalisation de la taille

La taille des caractères peut varier d'une fonte à l'autre et même au sein d'une même fonte après agrandissement ou réduction. La normalisation de la taille permet de ramener les images de mots à la même taille.

Figure 1.8. Résultats de la normalisation [14].

Segmentation

La séparation des lignes, des mots, des pseudo-mots, des caractères constituent l'opération de segmentation de l'écriture manuscrite.

a) **Segmentation d'un bloc de texte en lignes** : Le but de cette étape est la séparation des différentes lignes du texte en s'appuyant souvent sur les histogrammes de projection horizontale. Cependant, le chevauchement et l'interférence des ascendants et des descendants des lignes voisines rendent difficile l'opération de segmentation [14].

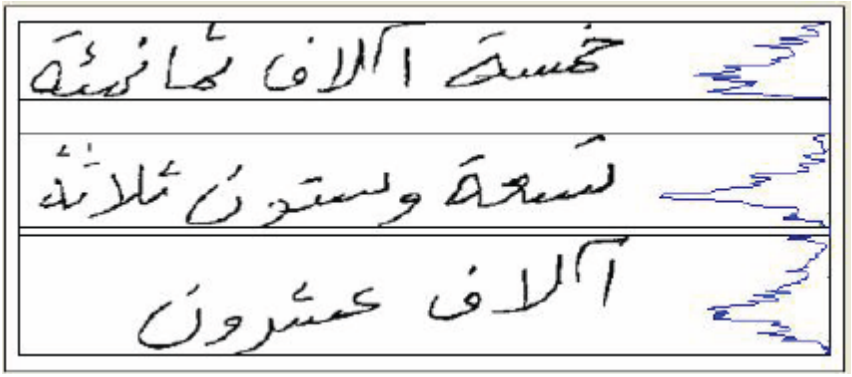

Figure 1.9. Exemple de localisation des lignes d'écriture [14].

b) **Segmentation des lignes en mots ou en sous-mots (pseudo-mots) :** elle se fait par la détection des espaces entre les mots et les pseudo-mots en utilisant les projections verticales,

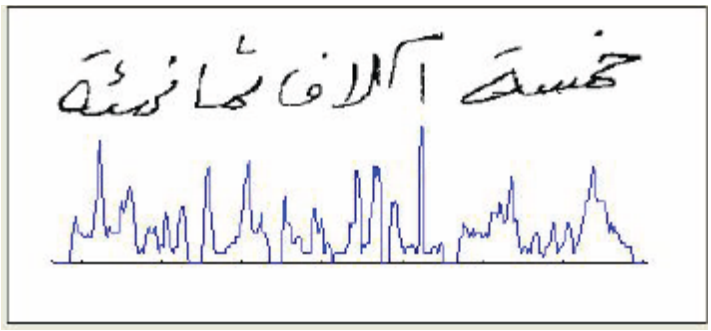

Figure 1.10. Histogramme de la segmentation mots [14].

c) **Segmentation des mots en caractères ou graphème :** La segmentation en caractères (ou en parties de caractères), constitue le problème le plus délicat de la reconnaissance de l'écriture.

1.8.3. Extraction des caractéristiques

Au lieu d'alimenter directement le module de reconnaissance avec l'image du caractère, il est préférable de la paramétrer grâce à l'extraction de certaines de ses caractéristiques.

Cette étape de la reconnaissance consiste à extraire des caractéristiques permettant de décrire de façon non équivoque les formes appartenant à une même classe de caractères tout en les différenciant des autres classes.

Les primitives sont classées en deux catégories [4] : les primitives structurelles (ou primitives locales) basées sur une représentation linéaire du caractère, ladécomposition du caractère en segments de droites et courbes, contours du caractère, squelette) et les primitives

statistiques (ou primitives globales) basées sur la distribution des points, les transformations et les mesures physiques faites sur le caractère (surface, moments, projections). Ces deux types peuvent être combinés formant un vecteur de primitives.

1.8.4. Classification

L'étape de classification se fait en deux opérations ; l'apprentissage et la reconnaissance. L'apprentissage permet de construire un dictionnaire de prototypes. Il s'agit de regrouper plusieurs prototypes dont les primitives se rapprochent en classes.

Le processus de reconnaissance peut toujours se résumer à une décision de classification. Les réseaux de neurones constituent l'outil le plus utilisé dans les systèmes de reconnaissance grâce à leur capacité d'apprentissage. A partir de l'étape d'apprentissage, le système doit ajuster ses paramètres afin de donner une réponse lors de la phase de reconnaissance.

Toutes ces opérations sont extrêmement liées les unes avec les autres. La réussite de toute la chaine est conditionnée par une réussite de chacune de ses parties.

En plus de ces quatre phases, on pourra trouver une phase de post-traitement qu'est une opération facultative et dont le rôle consiste à vérifier et éventuellement à corriger l'hypothèse de reconnaissance générée par le module de classification.

1.9. Conclusion

Dans ce chapitre, après avoir introduit quelques aspects de la reconnaissance automatique de l'écriture, nous avons présenté un aperçu des différentes techniques employées dans le domaine. Nous avons dressé les différentes étapes d'un système de reconnaissance de caractères et de mot à savoir les prétraitements, l'extraction des caractéristiques et les méthodes de classification. Nous nous focaliserons dans le chapitre suivant sur l'extraction des caractéristiques et la méthode de classification par les réseaux de neurones multicouches.

Chapitre 2

Extractions des caractéristiques et classification

2.1. Introduction

Dans ce chapitre, nous décrivons, tout d'abord, l'étape d'extraction des caractéristiques qui consiste à l'obtention de l'information la plus pertinente d'une forme, pour la présenter au système de reconnaissance. Les caractéristiques ou primitives d'une forme constituent une représentation réduite de la forme d'où une perte irrémédiable d'information s'impose.

Dans la deuxième partie de ce chapitre, nous présentons les réseaux de neurones artificiels (RNA) en rappelant le modèle formel. Ensuite, nous donnons une description de l'architecture générale d'un réseau de neurones et de son modèle de fonctionnement, et particulièrement pour un perceptron et un réseau multicouches. Nous terminons cette partie par une revue des exemples d'utilisation des perceptrons multicouches dans la reconnaissance automatique de l'écriture.

2.2. Extraction des caractéristiques

L'extraction des paramètres caractéristiques est une étape de grande importance. Si elle est mal conçue, il sera difficile, voire impossible, d'effectuer une reconnaissance efficace. Nous allons caractériser la forme des chiffres par des paramètres que nous voyons pertinents pour leur forme. Pour cela, nous allons détecter sur le chiffre des zones caractéristiques :les différents types de cavités, les histogrammes de projection, les moments géométriques, et les moments Zernike.

2.2.1. Approche VH2D (vertical-horisontal-2diagonal)

L'approche VH2D consiste à projeter chaque caractère selon les quatre directions ; l'abscisse, l'ordonnée et les deux diagonales 45° et 135°.Les projections s'effectuent en calculant la somme des valeurs des pixelsIxyselon la direction donnée [13].

Projection horizontale

$$p_x^h = \sum_{y=1}^{N} I_{xy} \quad , \quad P^h = \left[p_1^h, p_2^h, \dots p_N^h\right] \tag{2.1}$$

Projection verticale

$$p_y^v = \sum_{x=1}^{N} I_{xy} \quad , \quad P^v = [p_1^v, p_2^v, \dots p_N^v] \tag{2.2}$$

Projection sur la diagonal 45°

$$p_m^{d1} = \begin{cases} \sum_{l=N-m+1}^{N} \sum_{k=1}^{m} I_{lk} & si\ 1 \le m \le N\ et\ l = k + N - m \\ \sum_{l=1}^{2N-m} \sum_{k=m-N+1}^{N} I_{lk} & si\ N+1 \le m \le 2N-1\ et\ l = k + N - m \end{cases}$$

$$P^{d1} = \left[p_1^{d1}, p_2^{d1}, \dots p_{2N-1}^{d1}\right] \tag{2.3}$$

Projection sur la diagonal 135°

$$p_m^{d2} = \begin{cases} \sum_{l=1}^{m} \sum_{k=1}^{m} I_{lk} & si\ 1 \le m \le N\ et\ k = m - l + 1 \\ \sum_{l=1}^{2N-m} \sum_{k=m-N+1}^{N} I_{lk} & si\ N+1 \le m \le 2N\ et\ k = m - l + 1 \end{cases}$$

$$P^{d2} = \left[p_1^{d2}, p_2^{d2}, \dots p_{2N-1}^{d2}\right] \tag{2.4}$$

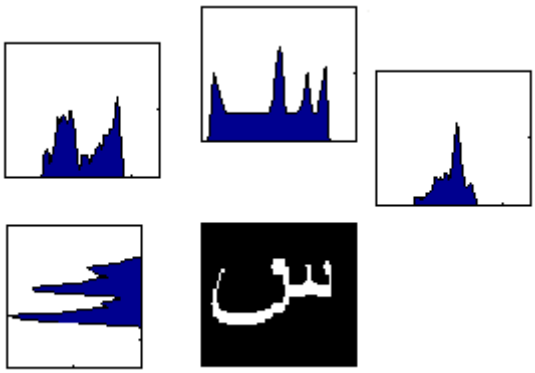

Figure 2.1. Exemple de projection

2.2.2. Zonage

Tout caractère s'identifie par la répartition des pixels qui le constitue. Le zonage consiste à diviser l'image en plusieurs zones de différentes manières. Les répartitions des pixels du caractère dans chaque zone constituent les composantes du vecteur des caractéristiques [13].

$$R_j = \frac{n_j}{n} \tag{2.5}$$

n_j, le nombre de pixels noirs du caractère dans la région $j\ et\ \ n\ est$ le nombre total de pixels noirs du caractère.

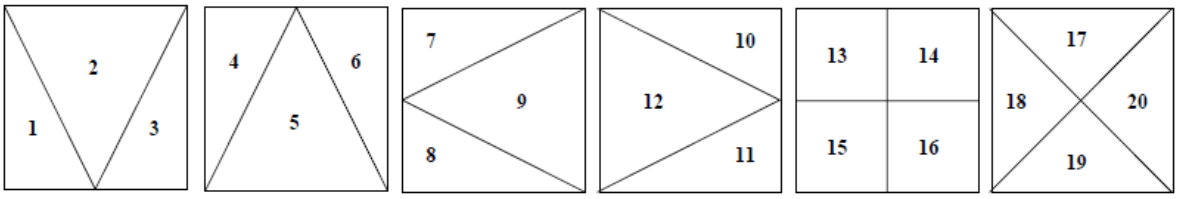

Figure 2.2. Différentes zones

2.2.3. Profils

Les profils décrivent bien les formes externes des caractères. La mesure de ces profils consiste à calculer le nombre de pixels (distance) entre le bord de l'image et le premier pixel du caractère rencontré. Pour discriminer efficacement les caractères, plusieurs profils peuvent être utilisés, le profil gauche, droite, supérieur, inferieur et orienté.

2.2.4. Cavités

Ces caractéristiques représentent les cavités dans un chiffre. Ces cavités ont été introduites par Michelle pour définir un modèle symbolique du chiffre et sont localisées grâce à la notion de relaxation morphologique développée par komri [13].

On définit 5 types de cavités : Ouest, Nord, Sud, Est, et Centre ,Cela suppose la définition de 4 directions cardinales dans l'image : Ouest vers la gauche, Nord vers le haut, sud vers le bas et l'est vers la droite. Par exemple, un point de l'image appartient à une cavité Ouest si et seulement si les conditions suivants sont simultanément vérifiées ;

- Ce point n'appartient pas au tracé.
- A partir de ce point, en se déplaçant en ligne droite vers l'Ouest, on ne croisl pas le tracé.
- A partir de ce point, se déplaçant en ligne droite vers le Nord, Sud, ou l'Est, on croise le tracé.

Par analogique, on peut facilement en déduire les définitions des cavités Nord, Sud, Est, Centre. Ces paramètres seront présentés au réseau sous la forme d'un rapport de leurs surfaces totales des cavités du chiffre donné.

La figure 2.3 montre les cavités détectées sur les chiffres. Chaque type de cavité est représenté par une couleur (Est : rouge, Ouest : vert, Sud : bleu, Centrale : violet).

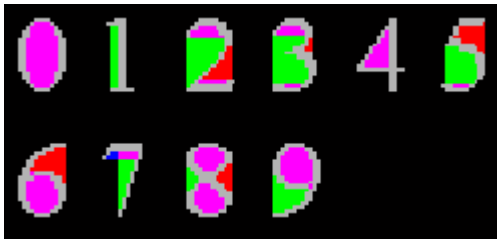

Figure 2.3. Différentes cavités.

2.2.5. Moments géométriques

La méthode des moments géométriques, comme toutes les méthodes statistiques, permet d'extraire des paramètres propres à la forme à reconnaître. Hu a élaboré un théorème énonçant

qu'un ensemble infini de moments calculés à partir d'une image permet de la décrire uniquement et complètement. Il a aussi créé une version de ces moments invariant E en translation, en rotation et en homothétie [8].

Les moments d'ordre $p+q$ d'une fonction $f(x,y)$sont définis par :

$$m_{pq} = \iint_{-\infty}^{+\infty} x^p\, y^q f(x,y)dxdy \tag{2.6}$$

Où p et q sont l'ordre du moment par rapport à x et y respectivement. Dans le cas d'une image binaire où :

$$f(x,y) = \begin{cases} 1 & (x,y) \in objet \\ 0 & sinon \end{cases} \tag{2.7}$$

L'équation (2.6) devient :

$$m_{pq} = \sum_x \sum_y x^p\, y^q \tag{2.8}$$

Pour rendre les moments invariants par rapport à la translation, l'image doit être centrée par rapport à son centre de gravité. Le moment central résultant aura la forme :

$$\mu_{pq} = \sum_x \sum_y (x - x_c)^p\, (y - y_c)^q \tag{2.9}$$

où x_c et y_c sont les cordonnés du centre de gravité de l'image de l'objet données par :

$x_c = \dfrac{m_{10}}{m_{00}}$, $y_c = \dfrac{m_{01}}{m_{00}}$

$\mu_{00} = \mathrm{m}_{00}$

$\mu_{10} = \mu_{01} = 0$

$\mu_{11} = m_{11} - (m_{00}x_c y_c)$

$\mu_{20} = m_{20} - (m_{00}x_c^2)$

$\mu_{02} = m_{02} - (m_{00}y_c^2)$

$\mu_{21} = m_{21} - (m_{20}y_c) - (2m_{11}x_c) - (2m_{00}x_c^2 y_c)$

$\mu_{12} = m_{12} - (m_{02}x_c) - (2m_{11}y_c) - (2m_{00}y_c^2 x)$

$\mu_{30} = m_{30} - (3m_{20}x_c) + (2m_{00}x_c^3)$

$\mu_{03} = m_{03} - (3m_{20}y_c) + (2m_{00}y_c^3)$

Les moment centrés et normalisés η_{pq} sont obtenus en divisant chaque moment centré μ_{pq} par μ^{γ}:

$$n_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \quad \gamma = \frac{p+q}{2} + 1 \tag{2.10}$$

Ces moments sont caractérisés par une invraiance par rapport à la translation et le changement d'echelle.

Les moments de Hu sont donnés par :

$\emptyset_1 = \eta_{20} + \eta_{02}$

$\emptyset_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$

$\emptyset_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$

$\emptyset_4 = (\eta_{20} - \eta_{02})^2 + (\eta_{21} - \eta_{03})^2$

$\emptyset_5 = (\eta_{30} - 3\eta_{12})\,(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} - \eta_{03})^2) + (3\eta_{21} - \eta_{03}) + (3\eta_{30} - 3\eta_{12}) + (\eta_{21} - \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2)$

$\emptyset_6 = (\eta_{20} - \eta_{02})(\eta_{20} - \eta_{02})^2 - (3\eta_{21} - \eta_{03})^2) + \eta_{11}(\eta_{03} - \eta_{12})(\eta_{21} - \eta_{03})$

$\emptyset_7 = (\eta_{21} - \eta_{03})\,(\eta_{30} - 3\eta_{12})((\eta_{30} + \eta_{12})^2 - (3\eta_{21} - \eta_{03})^2) - (\eta_{30} + 3\eta_{12})\,(\eta_{21} - \eta_{03})(\,(3\eta_{21} - \eta_{03})((\eta_{30} + \eta_{12})^2 - (\eta_{21} - \eta_{03})^2)$

L'intérêt de l'utilisation des moments géometrique dans la reconnaissance des caractères, reside dans le fait que ces moments nous permettent de trouver des caracteristiques propres aux formes des caractères à reconnaitre. Le calcul de ces momoents se fait comme suit, après la lecture de l'image du caractère, on calcule ces moments m_{pq} , puis les moments centraux μ_{pq} et enfin les moments invariants $\emptyset_k$

Le tableau 2.1 montre les 7 moments de Hu du chiffre 5 en lui faisant subir un changement d'echelle, une rotation et une translation.

On remarque qu'une rotation de 90° ou une translation n'entraine aucune modification des moments. Cependant, on constate une légere modification pour un changement d'echelle ou une rotation de 45°.

Tableau 2.1. Moments de Hu du chiffre '5' avec differentes transformations

φ_1	0.5030	0.5036	0.5031	0.4993	0.5030	0.5030
φ_2	0.0633	0.0642	0.0634	0.0641	0.0633	0.0633
φ_3	0.0006	0.0005	0.0006	0.0005	0.0006	0.0006
φ_4	0.0005	0.0005	0.0005	0.0011	0.0005	0.0005
φ_5	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
φ_6	0.0001	0.0001	0.0001	0.0003	0.0001	0.0001
φ_7	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000

2.2.6. Moments de Zernike

Les moments de Zernike sont basés sur le principe des polynômes orthogonaux. De ce fait, la reconstruction (en général approchée) de la forme à partir de ces moments est possible. Un ensemble relativement petit de moments de Zernike peut caractériser la forme globale d'un objet. Les moments d'ordre bas représentent la forme globale d'un objet et celles d'ordre plus supérieur le détail de cet objet. Plus on fait appel à des moments de Zernike d'ordre plus élevé, plus est la précision de reconstruction de l'image de l'objet. Pour des images binaires, un ordre modeste est suffisant pour la reconstruction.

Les moments de Zernike sont qualifiés par leur invariance par translation, par changement d'échelle et par rotation [9,10].

Les polynômes de Zernike notés par $V_{nm}(x,y)$ sont donnés par :

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)e^{jm\theta} \tag{2.11}$$

Où n est' un entier positif ou nul, $m\ est\ un$ entier posiyif ou négatif selon n ; est ta ρlongueur du vecteur de l'origine au pixel (x,y) et θ est l'angle entre le vecteur ρ et l'axe des abscisses dans le sens des aiguille d'une montre.

R_{nm}est un polynôme radiale donné par :

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2}(-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}\rho^{n-2s} \tag{2.12}$$

Les moments de Zernike sont les projections de l'image sur la base des fonctions orthogonales $V_{nm}(\rho,\theta)$.

Les moments de Zernike d'ordre n avec une répétition m pour une fonction image $f(x,y)$ qui s'annule en dehors du cercle unitaire ont pour expression :

$$A_{nm} = \frac{n+1}{\pi}\iint_{x^2+y^2\leq 1} f(x,y)V_{nm}^*(\rho,\theta)dxdy \tag{2.13}$$

Pour une image numérique :

$$A_{nm} = \frac{n+1}{\pi}\sum_x\sum_y f(x,y)V_{nm}^*(\rho,\theta) \tag{2.14}$$

Pour calculer les moments de Zernike d'une image donnée, le centre de l'image est pris comme origine.

Pour ramener l'image à l'intérieur du cercle unitaire, on réalise une normalisation des rayons de tous les pixels de l'image : Partant du centre de gravité, on balaye tous les pixels de l'image pour chercher le plus grand rayon, celui-ci est considéré comme unitaire.

Tous les autres rayons sont divisés par ce rayon maximum pour inclure les pixels correspondants dans le cercle unité. Le tableau 2.2 montre les 12 premiers moments de

Zernike du chiffre 5 en lui faisant subir un changement d'echelle, une rotation et une translation.

Tableau 2.2. Moments de zernike du chiffre '5' avec differentes transformations

	Image ori	50%	200%	Rot 45	Rot 90	trans
A_{00}	0.2530	0.2530	0.2548	0.2538	0.2530	0.2530
A_{11}	0.0754	0.0755	0.0762	0.0708	0.0754	0.1434
A_{20}	0.1144	0.1131	0.1138	0.1216	0.1144	0.0651
A_{22}	0.1419	0.1430	0.1435	0.1437	0.1419	0.0806
A_{31}	0.1147	0.1175	0.1163	0.1261	0.1147	0.0444
A_{33}	0.0791	0.0791	0.0804	0.0736	0.0791	0.1367
A_{40}	0.0643	0.0580	0.0641	0.0708	0.0643	0.0667
A_{42}	0.0429	0.0383	0.0419	0.0534	0.0429	0.1766
A_{44}	0.0213	0.0228	0.0211	0.0188	0.0213	0.0672
A_{51}	0.1684	0.1607	0.1680	0.1623	0.1684	0.1309
A_{53}	0.1416	0.1469	0.1443	0.1370	0.1416	0.1571
A_{55}	0.0378	0.0405	0.0390	0.0267	0.0378	0.0073

2.3. Classification

La classification est l'étape de décision ; le choix de la classe dont la représentation ou le modèle est le plus proche. Il existe de nombreuses méthodes de classification, cependant on se limite dans ce qui suit au classificateur neuronal.

Les réseaux de neurones artificiels constituent une imitation du cerveau humain dans la résolution des problèmes que les êtres vivants savent résoudre. Ils se ressemblent en deux points essentiels ; la connaissance est acquise au travers d'un processus d'apprentissage, les poids des connections entre les neurones sont utilisés pour mémoriser la connaissance [1].

2.3.1. Neurone formel

Le modèle du neurone formel est inspiré du neurone biologique. Il peut être considéré comme un processeur élémentaire qui réalise une somme des signaux qui lui parviennent $x_i, i = 1, \dots, N$, pondérée par l'efficacité synaptique correspondante w_i . Cette quantité, en lui enlevant un seuil interne θ du neurone, constitue l'activation ou le potentiel du neurone [1]:

$$a = \sum_{i=1}^{N} x_i w_i - \theta \tag{2.15}$$

La sortie ou l'activité du neurone est obtenue par le passage du potentiel du neurone à travers une fonction appelée une fonction d'activation.

$$y = f(\sum_{i=1}^{N} x_i w_i - \theta) \tag{2.16}$$

Le neurone formel est l'élément de base d'un réseau de neurones artificiel.

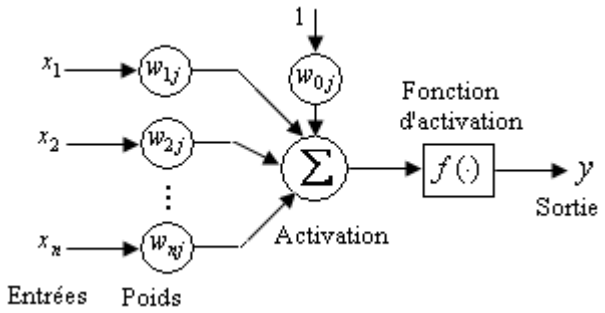

Figure 2.4. Modèle de neurone formel.

2.3.2. Perceptron

Le perceptron est le premier modèle des réseaux de neurones, il se compose de deux couches de neurones, la rétine (non comptable d'où le nom monocouche) et la couche de sortie. La sortie du neurone i est de la forme [1]:

$$y_i = f\left(\sum_{j=1}^{N} x_i w_{ij}\right) \tag{2.17}$$

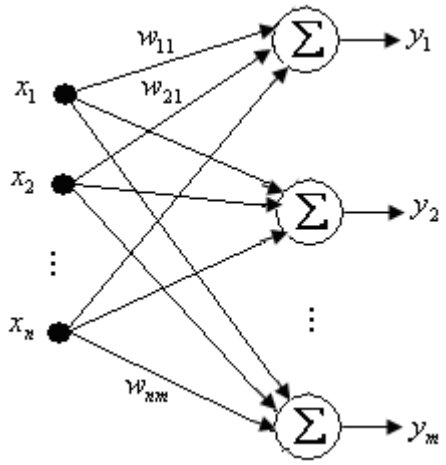

Figure 2.5. Le perceptron.

2.3.3. Perceptron multicouches

Le perceptron multicouches est une généralisation du perceptron, il se compose d'une couche d'entrée (rétine du perceptron), d'une ou plusieurs couches intermédiaires (dites

couches cachées), et d'une couche de sortie. Le calcul de l'activation dans le réseau s'effectue en propageant l'activation initiale de la couche d'entrée jusqu'à la couche de sortie [7].

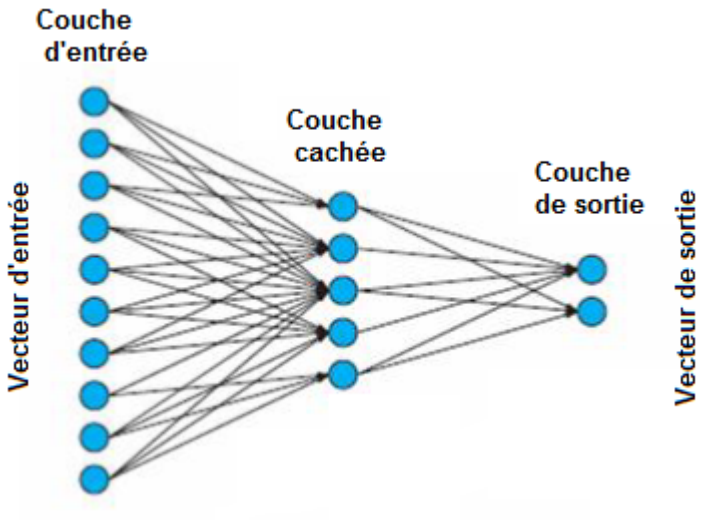

Figure 2.6. Perceptron multicouche fortement connecté aux couches cachées.

2.3.4. Règles d'apprentissage

L'apprentissage consiste à déterminer ou ajuster les poids synaptiques de façon à minimiser un critère d'erreur entre les vecteurs d'entrée et de la sortie désirée correspondante. On présente les vecteurs d'entrée au réseau après l'initialisation des poids synaptiques d'une façon aléatoire, ensuite on calcule la réponse du réseau. L'erreur engendrée doit être minimisée pour obtenir les poids synaptiques permettant au réseau d'apprendre les exemples d'apprentissage. L'algorithme utilisé est basé sur la minimisation de l'erreur quadratique moyenne. C'est la base de la rétro-propagation de l'erreur [1].

2.3.5. Rétro-propagation du gradient

L'algorithme de rétro-propagation de l'erreur consiste à minimiser l'erreur totale E_t qui est définie comme la somme des erreurs partielles obtenues pour chaque exemple du vecteur d'entrée.

$$E_T = \sum_p E_p \tag{2.18}$$

Avec E_pl'erreur de la p^{me}entrée de l'ensemble d'apprentissage qui est défini sous la forme :

$$E_p = \frac{1}{2}\sum_{j=1}^{M}(d_{pj} - z_{pj})^2 \tag{2.19}$$

où M représente le nombre de sortie du réseau, d_{pj} la sortie désirée et z_{pj} la sortie calculée.

Par conséquent, pour que la fonction coût E_p soit petite il est raisonnable de modifier ou faire varier les paramètres dans le sens négatif du gradient de E_p. Le principe de l'entraînement est de présenter le vecteur d'entrée de chaque exemple d'apprentissage et de calculer la réponse du réseau après l'avoir initialiser au hasard (poids). On calcule le gradient de l'erreur par rapport à tous les poids, puis les paramètres seront ajustés dans la direction opposée à celle du gradient de l'erreur. Cette opération est cyclique jusqu'à ce que l'erreur tant vers zéro.

2.3.6. Convergence et adaptation des poids

La mise à jour des poids synaptiques w_{ij}, est faite par la minimisation de l'erreur E_p. Le développement de l'expression de la dérivée partielle de E_p par rapport à w_{ij} nous donne:

a- Pour les neurones de sortie:

$$\frac{\delta E_p}{\delta w_{ij}^{\ l}} = (d_{P_j} - z_{P_j}) f^{l'}(Y_{Pj}^{\ l}) X_{Pi}^{\ l} \tag{2.20}$$

b-Pour les neurones des couches cachées :

$$\frac{\delta E_p}{\delta w_{ij}^{\ l}} = [\sum_K \frac{\delta E_p}{\delta Z_{PK}^{l+1}} f^{(l+1)'}(Y_{Pk}^{\ l+1}) w_{kj}^{\ l+1}] f^{l'}(Y_{Pj}^{\ l}) X_{Pi}^{\ l} \tag{2.21}$$

La formule itérative pour mesurer les paramètres d'un réseau est la suivante :

$$w^l_{\ ij}(t+1) = w^l_{\ ij}(t) + \Delta p w^l_{\ ij}(t+1) \tag{2.22}$$

Avec: $$\Delta p w^l_{\ ij}(t+1) = -\mu \frac{\delta E_p}{\delta w_{ij}^{\ l}(t)} \tag{2.23}$$

μ: est facteur d'apprentissage; leurs valeurs entre 0 et 1. La formule (2.22) s'écrit encore sous la forme :

$$w^l_{ij}(t+1) = w^l_{ij}(t) + \mu \delta^l_{Pj} X^l_{Pi} \tag{2.24}$$

Où δ^l_{Pj} est l'erreur équivalent à la sortie de neurone j de la couche l telle que :

$$\delta^l_{Pj} = -\frac{\partial E_p}{\partial Y^l_{Pj}} \tag{2.25}$$

Pour minimiser l'erreur totale E_t, on remplace l'équation (2.23) par :

$$\Delta w^l_{ij}(t+1) = -\mu \frac{\partial E_T}{\partial w^l_{ij}(t)} \tag{2.26}$$

La vitesse de convergence dépend de μ, si μ est trop petit la convergence de l'algorithme demande un large nombre d'itération, sinon, soit la convergence est rapide, soit il ne converge

pas à cause de la présence d'un phénomène d'oscillation qui intervient dès qu'on se rapproche de minimum. On ajoute un moment $\propto \in [0,1]$ à l'algorithme de rétro-propagation pour éviter leur problème d'être très lente à la convergence, et la loi d'adaptation devient :

$$w^{l}_{ij}(t+1) = w^{l}_{ij}(t) + \mu\delta^{l}_{Pj} X_P + \alpha(w^{l}_{ij}(t) - w^{l}_{ij}(t-1))\, p \qquad (2.27)$$

2.3.7. Algorithme de rétro-propagation

Les étapes de l'algorithme de rétro propagation sont résumées dans ce qui suit [7] :

1. Initialiser les poids avec des petites valeurs aléatoires (w_{ij} entre 0 et 1).
2. Présenter le vecteur d'entrée $X_p$de l'exemple p avec la sortie désirée D_p
3. Calculer le vecteur de sortie actuelle E_p et l'erreur correspondante E_p en utilisant les équations (2.18) et (2.19).
4. Calculer les dérivées partielles de l'erreur par rapport à chaque poids à l'aide des équations suivantes :
5. Si j est le neurone de la couche de sortie :
6. $$\delta^{l}_{pj} = f^{l'}(Y^{l}_{pj})(d_{pj} - z_{pj}) \qquad (2.28)$$
7. Si j est la neurone de la couche cache :
8. $$\delta^{l}_{pj} = f^{l'}(Y^{l}_{pj})\sum_{k=1}^{N_{L+1}} \delta^{l+1}_{pk} w^{l+1}_{kj} \qquad (2.29)$$
9. Ajuster les poids w_{ij} en utilisant l'équation (2.24) ou l'équation (2.27)
10. Répéter les étapes 2 à 6 pour l'ensemble des observations de la base d'apprentissage tant que le critère d'arrêt n'a pas été atteint.

2.3.8. Critère d'arrêt

Il existe plusieurs critères d'arrêts. Ces critères peuvent être combinés entre eux. Le premier critère est basé sur l'amplitude du gradient de la fonction d'activation, puisque par définition le gradient sera à zéro au minimum. Le second critère d'arrêt est de fixer un seuil que l'erreur quadratique ne doit pas dépasser. Toutefois ceci exige une connaissance préalable de la valeur minimale de l'erreur qui n'est pas toujours disponible. Le troisième critère consiste à déterminer un nombre fixe de cycles à atteindre [3].

2.4. Application des réseaux de neurones à la reconnaissance de l'écriture

Les perceptrons multicouches sont fréquemment utilisés dans les systèmes de reconnaissance automatique des textes. En effet, plusieurs systèmes à base des perceptrons

multicouches sont proposés pour la reconnaissance des caractères manuscrits. Parmi les applications des réseaux de neurones, on cite à titre d'exemples la lecture des montants littéraux et numériques de chèques, la lecture du code postal dans le tri des lettres [5,7,13].

2.5. Conclusion

L'objectif de ce chapitre est de donner un aperçu général sur les deux dernières étapes du système de reconnaissance de caractère à savoir l'extraction des caractéristiques et la classification. Pour les caractéristiques, on a présenté les caractéristiques les plus représentatives comme les cavités, les projections, le zonage, le profil, les moments Hu et les moments Zernike. Les quatre premiers caractéristiques sont très sensibles à la rotation ou à la translation. Par contre, les moments de Hu et de Zernike sont invariants à la translation, au grossissement, et à la rotation. Pour la classification, on s'est limité aux perceptrons multicouches.

Chapitre 3

Application à la reconnaissance de caractères arabes

3.1. Introduction

Dans ce chapitre, on va présenter les résultats de classification des chiffres et des mots arabes par le classifieur neuronal. Plusieurs structures de réseaux de neurones ; nombre de couches cachées, nombre de neurones dans chaque couche cachée seront présenté. Les caractéristiques représentant les caractères ou les mots seront utilisées séparément dans un premier lieu puis. Pour mieux augmenter le taux de reconnaissance, les caractéristiques seront combinées de différentes.

3.2. Reconnaissance des chiffres arabes

La première phase de notre travail se focalise sur la reconnaissance hors ligne des chiffres arabes manuscrits utilisant un perceptron multicouches. L'objectif est de déterminer les caractéristiques les plus discriminantes, ainsi que l'architecture du réseau MLP.

3.2.1. Système de reconnaissance proposé

Le processus de reconnaissance proposé se déroule en trois étapes :

Présentation des chiffres arabes

Les chiffres utilisés sont issus de la base MNIST [11]. C'est une banque de données d'images de chiffres manuscrits initiée par Y. Lecun et C. Cortes afin de produire un grand nombre de données réelles et variées dans le domaine de l'apprentissage et de la reconnaissance de lettres manuscrites. Elle contient un ensemble de 60 000 images utilisées pour l'apprentissage et un ensemble de 10 000 images pour les tests. Toutes les images sont en niveaux de gris et de taille 28x28 pixels. L'intérêt d'utiliser MNIST est que les images ont toutes la même taille et ont été normalisées et centrées, ce qui évite avantageusement les étapes de prétraitement. La base MNIST est disponible publiquement et gratuitement à cette adresse http://yann.lecun.com/exdb/mnist/. La figure 3.1 montre des images de dix chiffres issus de la base MNIST.

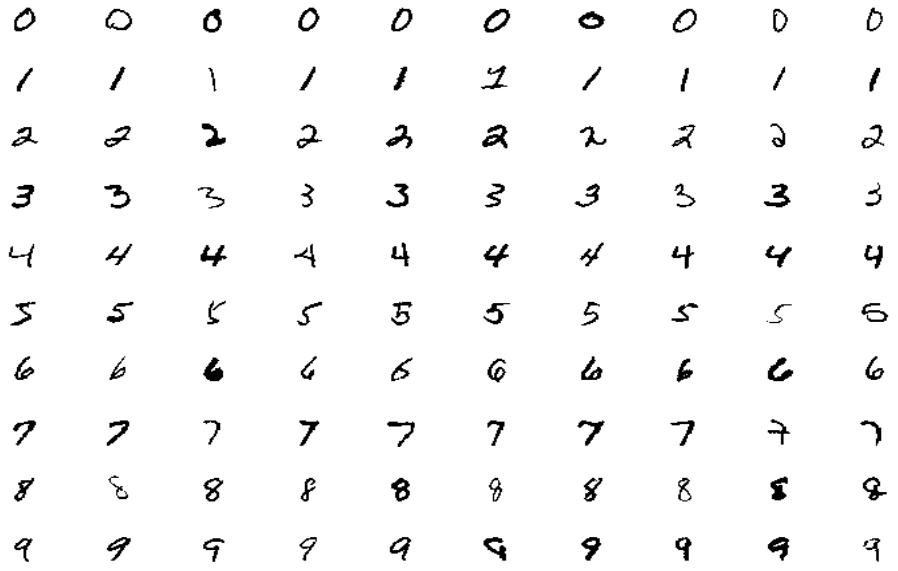

Figure 3.1. Exemples d'images extraites de MNIST.

Nous n'avons utilisé que 10000 images provenant de la base d'apprentissage pour l'apprentissage et 1000 images de la base de test pour la validation. Aucun traitement supplémentaire n'a été réalisé sur ces images.

Extraction des caractéristiques

L'objectif de l'extraction des caractéristiques consiste à exprimer les caractéristiques sous une forme numérique. La difficulté ici est de trouver de bonnes caractéristiques permettant aux classifieurs de reconnaître facilement les différents chiffres. Les caractéristiques utilisées sont :

- HV2D : histogrammes des projections horizontales, verticales et diagonales (45°, 135°) du chiffre.
- Ca : Cavités sud, est, ouest, nord et centre du chiffre.
- Hu : Moments de Hu d'ordres 7.
- Ze : Moments de Zernike d'ordre 9.

On obtient un vecteur de caractéristiques qui regroupe 180 caractéristiques réparties comme suit :

- 5 caractéristiques de type concavité par chaque chiffre.
- 138 caractéristiques de type projection par chaque chiffre.
- 7 caractéristiques des moments de Hu
- 30 caractéristiques des moments de Zernike

Classification

Le classifieur utilisé dans ce travail est un réseau de neurones de type MLP à plusieurs couches cachées. Le nombre de neurones dans les couches cachées est déterminé à partir de plusieurs expériences. Le nombre d'entrées est fixé selon les caractéristiques utilisées. La couche de sortie contient 10 neurones correspondant aux 10 classes des chiffres.

L'apprentissage

Chaque image des chiffres de la base d'apprentissage est étiquetée. Cette étiquette détermine la classe du chiffre présenté. Donc pour chaque image présentée au système de reconnaissance, on doit attribuer son étiquette à la sortie du réseau. De cette façon, le réseau doit ajuster ses poids pour ressembler les chiffres d'une même classe et dissembler les chiffres de classes différentes. Une fois l'apprentissage effectué, le système doit être validé par une phase de test.

Validation

L'étape de test permet d'évaluer la performance du classifieur pour un apprentissage donné. C'est une étape importante car elle peut mettre en cause le choix des caractéristiques ou la structure du réseau. En effet, il est difficile de trouver a priori les caractéristiques pertinentes et la structure la plus adaptée au problème posé. Si les performances du système souhaitées ne sont pas atteintes, il suffit de trouver une nouvelle famille de caractéristiques ou d'essayer une autre structure [3].

Critères d'évaluation

Pour une entrée donnée, un classifieur peut générer les réponses suivantes : Un rejet pour indiquer que le classifieur n'a pas pu identifier cette entrée, Une reconnaissance pour une bonne identification de la classe ou une confusion dans le cas où le classifieur attribue une autre classe à l'entrée. Pour évaluer la performance des réseaux conçus, nous allons utiliser les critères d'évaluation suivants :

- Taux d'apprentissage: il représente le taux de caractères correctement classés de la base d'apprentissage
- Taux de reconnaissance (ou de réussite) : le taux de caractères correctement classées de la base de test.
- Taux de confusion: nombre de caractères mal classés sur le nombre total des caractères de la base de test.
- Taux de rejets: nombre de caractères pour lesquelles le réseau ne donne pas une décision sur le nombre total des caractères de la base de test.

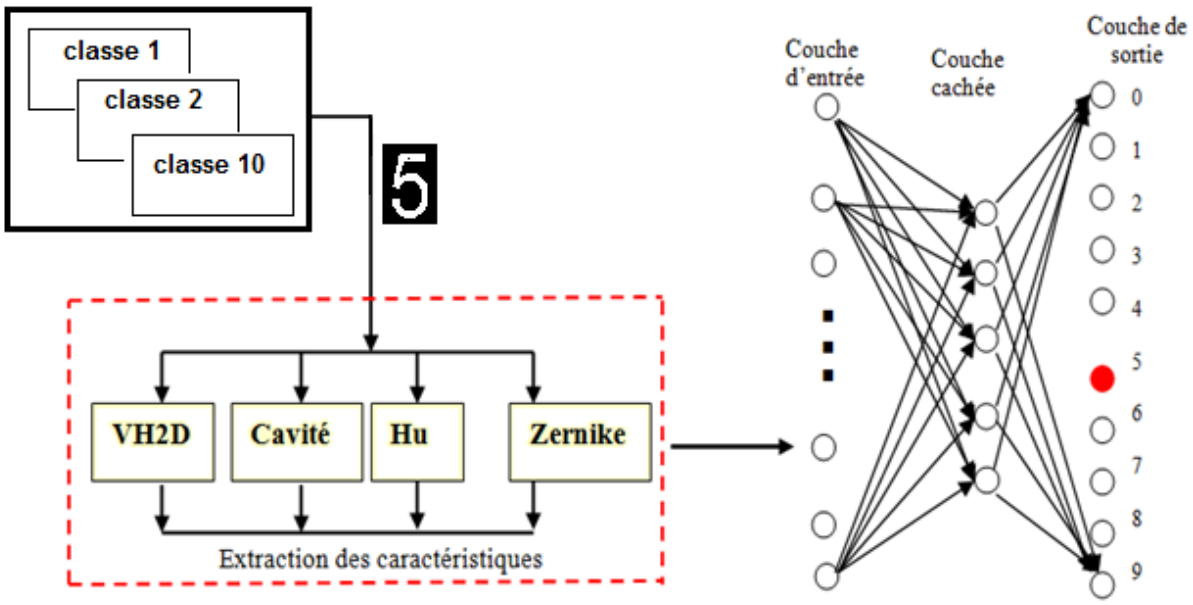

Figure 3.2. Système de reconnaissance utilisé.

3.2.2. Résultats de simulation

La détermination de la structure du réseau n'est pas une tache triviale, en effet la structure donnant les meilleures performances n'est obtenue qu'après une série d'expérimentation. Pour cela, on a considéré différentes architectures avec les différentes caractéristiques comme le montre le tableau 3.1**.**

D'après les résultats, on constate que les projections et les moments de Zernike offrent les meilleures performances de point de vue apprentissage et reconnaissance par rapport aux moments de Hu et les cavités. On note ici que les taux obtenus sont relativement faibles. Ceci est dû au nombre insuffisant d'itération pour l'apprentissage et ça ne doit pas donc pénaliser l'aptitude des réseaux de neurones dans la classification.

Tableau 3.1 Taux d'apprentissage et de reconnaissance pour la base MNIST. (Ta : Taux d'apprentissage, Tr : Taux de reconnaissance)

MLP	**Primitive**	**HV2D**		**Cavité**		**Hu**		**Zernike**	
	Performance	**Ta**	**Tr**	**Ta**	**Tr**	**Ta**	**Tr**	**Ta**	**Tr**
1 couche	20 neurones	94.01	86.46	67.46	64.24	49.31	49.00	79.17	74.80
	50 neurones	93.43	89.66	68.37	65.10	51.40	51.50	66.59	63.68
2 couches	25x25	94.61	85.80	73.57	68.78	54.52	54.84	83.2	77.06
	50x100	95.84	90.32	70.56	66.44	56.18	55.92	77.45	74.06
	100x100	85.33	82.52	71.41	67.22	51.05	50.32	83.71	78.98

Combinaison des caractéristiques

Pour augmenter la performance du système de reconnaissance, il est utile de combiner les différentes caractéristiques. Vu le temps nécessaires pour l'apprentissage, on n'a pas pu essayer toutes les combinaisons possibles. La combinaison qui regroupe les quatre caractéristiques a été adoptée. Le tableau 3.2. montre les résultats obtenus avec un réseau de deux couches cachées ayant chacune 100 neurones.

Tableau 3.2 Taux d'apprentissage et de reconnaissance avec combinaisons de caractéristiques.

2 couches cachées100x100	**Ta**	**Tr**	**Erreur résiduelle**
HV_Cavité_Hu_Zernike	**97.81**	**93.16**	**0.0038**
HV_Zernike	**91.96**	**88.96**	**0.0194**

Le tableau 3.3 montre la matrice de confusion pour la reconnaissance, Les colonnes correspondent aux valeurs reconnues, les lignes aux valeurs vraies.

Tableau 3.3 Matrice de confusion de l'apprentissage.

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Tr**
0	994	0	0	1	0	0	1	0	3	1	99.40
1	0	983	6	1	0	2	2	1	1	4	98.30
2	0	1	978	7	0	1	2	0	7	4	97.80
3	0	1	5	972	0	3	1	7	3	8	97.20
4	2	0	3	0	979	0	3	0	3	10	97.90
5	1	1	1	5	0	983	6	1	1	1	98.30
6	3	1	5	0	4	3	981	0	3	0	98.10
7	0	0	4	1	1	0	0	979	1	14	97.90
8	1	0	4	3	2	6	1	3	973	7	97.30
9	5	2	4	6	7	2	0	6	9	959	95.90
Ta : 97.81%						Tr **: 93.16%**					

L'analyse de la matrice de confusion de la reconnaissance permet de tirer les remarques suivantes :

- Les chiffres les plus connues sont '1', '0' , '8' et '4' par ordre décroissant.
- Les chiffres '9' et '7' sont les moins reconnus.
- une confusion du chiffre '9' avec '3', '4' ,'7' et '8' .
- une confusion du chiffre '7' avec '2', '9' et '3'

Tableau 3.4 Matrice de confusion de la reconnaissance

	0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Tr**
0	490	0	1	0	0	3	0	0	5	1	98.00
1	0	491	3	1	0	2	3	0	0	0	98.20
2	2	0	460	13	1	4	5	9	5	1	92.00
3	0	0	9	464	0	7	2	11	4	3	92.80
4	1	0	3	0	471	0	5	0	1	19	94.20
5	3	0	2	12	1	459	6	5	8	4	91.80
6	10	3	6	0	5	6	461	0	9	0	92.20
7	1	5	16	7	9	1	0	444	2	15	88.80
8	5	0	3	6	1	1	2	5	474	3	94.80
9	1	3	2	11	12	4	0	12	11	444	88.80
Ta : 97.81%						Tr **: 93.16%**					

Ces chiffres sont parfois visuellement très proches où la décision est difficile, même pour un opérateur humain.

3.3. Reconnaissance de mots arabes isolés à vocabulaire limitée

La deuxième phase du travail concerne la reconnaissance des mots isolés. L'approche globale a été adoptée pour la reconnaissance, c'est-à-dire que le réseau apprend les mots sans segmentation. Les mots utilisés sont issus du répertoire wrd-mst de la base AHDB fut proposée par S. Al-Ma'adeed [2]. C'est un ensemble de données qui été recueilli dans l'Université du Qatar et est essentiellement destiné à des tâches de reconnaissance de l'écriture manuscrite arabe. Elle contient des mots arabes écrits par une centaine de scripteurs. On note ici que les images sont binaires et non normalisés.

Figure 3.3 Exemples d'images extraites de la base AHDB

3.3.1. Caractéristiques utilisées

Du fait que les images sont non normalisées, la tache primaire consiste à les faire normaliser. Hors comme on a vu au chapitre 2, les moments de Hu ou les moments de Zernike ne sont pas sensibles à la translation ou à un changement d'échelle, de même, les

cavités constituent un rapport entre la surface de la cavité et la surface globale de l'image donc peu sensibles à la taille de l'image, ceci nous a orienté à utiliser ces trois caractéristiques sans normalisation des images. La figure 4.3 montre les caractéristiques utilisées pour les dix mots 'abd'.

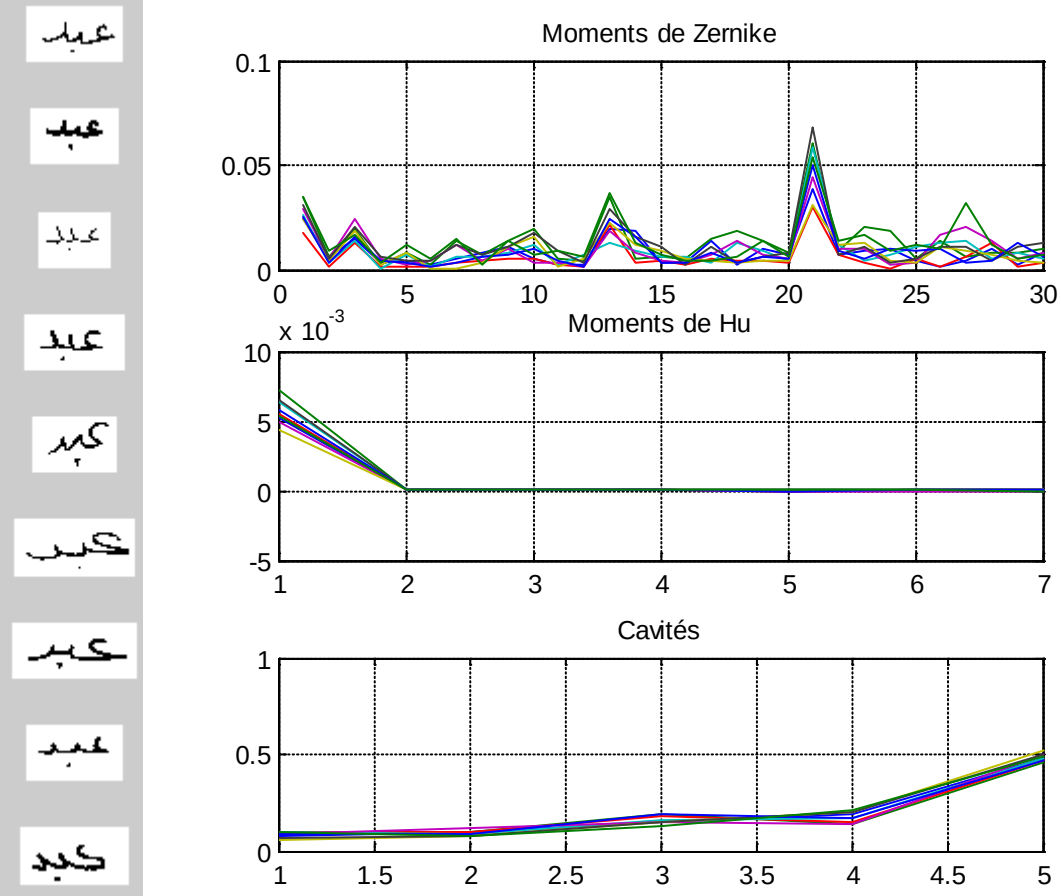

Figure 3.4 Caractéristiques utilisées

3.3.2. Résultats de simulation

Dans cette partie on a utilisé deux structures de réseau, le premier est constitué de deux couches cachées ayant 250, 500 neurones respectivement, une couche de sortie avec 28 neurones correspondant à la taille du vocabulaire. Le deuxième contient deux couches cachées ayant 300 neurones chacune.

Tableau 3.5 Taux d'apprentissage et de reconnaissance pour les mots isolés

Primitive	**Zernike**			**Cavité-Zernike-Hu**		
Performance	**Ta**	**Tr**	**Erreur d'app.**	**Ta**	**Tr**	**Erreur d'app.**
250x500	96.93	74.53	0.0133	97.70	79.40	0.0124
300x300	95.25	73.08	0.0154	90.40	75.18	0.0176

L'erreur moyenne quadratique minimale est obtenue avec le premier réseau. Le taux de reconnaissance a été augmenté avec la combinaison des trois caractéristiques.

3.4. conclusion

Dans ce chapitre nous avons présenté notre travail, commençant par une présentation de la base de mots et des chiffres utilisés, puis les caractéristiques choisies, ainsi que l'apprentissage et le test du classifieur neuronal. Enfin, la combinaison des caractéristiques nous a permis d'augmenter le taux de reconnaissance moyen.

Conclusion générale

La reconnaissance de l'écriture manuscrite joue un rôle très important dans différents domaines tels que la lecture automatique des chèques, des adresses postales, des formulaires administratifs. Quand il s'agit de l'écriture latine ou de l'écriture arabe imprimée, la reconnaissance a été effectuée avec un grand succès dû à la possibilité d'isoler les caractères par une étape de segmentation. Cette tâche est très difficile voire même impossible dans le cas de l'écriture arabe cursive ou semi-cursive. L'objectif de ce mémoire est de proposer un système de reconnaissance de l'écriture manuscrite arabe hors-ligne. Le travail réalisé porte essentiellement sur la reconnaissance des chiffres et les mots isolés. Pour les mots isolés, la technique globale de reconnaissance a été adoptée suite au vocabulaire limité utilisé.

Utiliser une image comme telle qu'elle c'est l'idéale, hors comme les mots sont des données de grandes dimensions, donc il faut extraire les caractéristiques les plus pertinentes de l'image pour ne garder qu'une quantité réduite d'informations. Plusieurs caractéristiques ont été utilisées. Chaque caractéristique possède des inconvénients et des avantages et à ce jour il n'existe pas de primitive fiable à 100 %. Pour remédier, il sera intéressant d'utiliser plusieurs primitives à la fois. Durant notre travail, quatre primitives ont été utilisées, les moments invariants de Zernike et de Hu, les cavités et les quatre projections HV2D.
D'après les résultats obtenus, on a constaté que les projections HV2D sont avantageuses pour la reconnaissance de chiffres. Hors pour les mots isolés, ce sont les moments invariants de Zernike qui donne de meilleurs résultats par rapport aux moments de Hu.

D'une manière générale, le but de ce projet n'est pas suffisamment atteint du fait que le taux de reconnaissance soit inférieur à 90% pour les mots isolés. Mais ceci n'est pas fatal, il s'explique par le nombre réduit d'exemples de la base d'apprentissage.

Pour achever, même si les réseaux de neurones sont caractérisés par une grande capacité d'apprentissage et de reconnaissance, rien ne garantit qu'un réseau de neurones conçu pour un problème de classification aura en toutes situations les performances attendues. La qualité de la base de données ainsi que les primitives sélectionnées influent énormément sur le système de reconnaissance des caractères manuscrits.

Bibliographie

[1] H. Abdi, *Les réseaux de neurones*. Press Universitaires de Grenoble, 1994.

[2] S. Al-Ma'adeed, D. Elliman, C.A. Higgins, 'A database for Arabic handwritten text recognition research', *in 8th International Workshop on Frontiers in Handwriting Recognition, Niagara-on-the-Lake,* Canada, 2002.

[3] M. Amrouch, 'Reconnaissance de caractères imprimés et manuscrits, textes et documents basée sur les modèles de Markov cachés', *Thèse de doctorat, Université Ibn Zohr Maroc*, 2014.

[4] A.Boukharouba , 'Contribution à la segmentation et à la reconnaissance de l'écriture arabe manuscrite', *Thèse de Doctorat, Université de Constantine* , 2011.

[5] M. Charfi, M. Kherallah, A. El Baati, and A.M. Alimi, 'A new approach for Arabic handwritten postal addresses recognition', *The International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 3, no. 3, pp. 1-6, 2012.

[6] L. Chergui, 'Combinaison de classifieurs pour la reconnaissance de mots arabes manuscrits', *Thèse de Doctorat, Université de Constantine,* 2013.

[7] B. Gosselin, 'Application de réseaux de neurones artificiels à la reconnaissance automatique de caractères manuscrits', *Thèse de Doctorat, Université de Belgique,* 1996.

[8] M. K. Hu, 'Visuel pattern recognition by moment invariants', *IRE Transactions On Information Theory,* pp. 179-187 , 1962

[9] C. Kan and M.D. Srinath, 'Invariant character recognition with Zernike and orthogonal Fourier-Mellin moments', *Pattern Recognition, Pergamon*, vol. 35, pp. 143-154, 2002.

[10] A. Khotanzad and Y. H. Hong , 'Invariant image recognition by Zernike moments', *IEEE Trans. Pattern Anal. Mach. Intell*. Vol.12 , pp.489–497, May 1990.

[11] Y. Lecun and C. Cortes , 'The MNIST database of handwritten digits',1990.

[12] F. Menasri, 'Contributions à la reconnaissance de l'écriture arabe manuscrite', *Thèse de doctorat, Université paris* , 2008.

[13] K. Merdour , 'Reconnaissance de formes du chiffre arabe imprimé : Application au code à barre d'un produit', *Thèse de Magister , Université de Batna* , 2009.

[14] A. Rebiai, 'Une approche hybride pour la reconnaissance d'écriture arabe manuscrite', *Thèse de Magister, Université de Constantine,* 2007.

Résumé

La reconnaissance de l'écriture manuscrite joue un rôle très important dans différents domaines tels que la lecture automatique des chèques, des adresses postales et des formulaires administratifs. L'objectif de ce projet est de proposer un système de reconnaissance hors-ligne des chiffres et mots arabes par les réseaux de neurones. Quatre types de caractéristiques des mots ont été utilisés, les moments invariants de Zernike et de Hu, les cavités et les quatre projections HV2D. Les données utilisées sont issues de deux bases de données, la base MNIST pour les chiffres et AHDB pour les mots isolés. Plusieurs architectures du réseau de neurones et primitives ont été testés pour avoir aboutir des taux de reconnaissance acceptables. Pour augmenter d'avantage le taux de reconnaissance, plusieurs combinaisons de caractéristiques ont été utilisées. Les résultats de simulation montrent que les moments de Zernike sont plus discriminants que les autres caractéristiques.

Mots-clés : Reconnaissance de l'écriture arabe manuscrite, Extraction des caractéristiques, Réseaux de neurones, Moments géométriques, Base de mots arabes.

Abstract

The handwriting recognition plays a very important role in various fields such as the automatic bank check, postal and administrative forms recognition. The objective of this project is to propose an off-line Arabic digits and words recognition system. Four kinds of features were used, the invariant Zernike and Hu moments, cavities and the four HV2D projections. The used data are provided from the MNIST database for digits and the AHDB for isolated words. Several neural network architectures and primitives were tested to lead to an acceptable recognition rate. To further increase the recognition rate, various combinations of primitives were used. Simulation results show that Zernike moments are more discriminating than other primitives.

Keywords: Arabic handwriting recognition, Feature extraction, Neural networks, Geometrical moments, Handwritten Arabic words databases.