

République Algérienne Démocratique et Populaire
Ministère de l'enseignement Supérieur et de la recherche scientifique



Université Mentouri de Constantine
Faculté des sciences de l'ingénieur

Département d'informatique

N° d'ordre :

N° série :

Mémoire

Pour l'obtention du diplôme de Magister en informatique
Option systèmes distribués

Thème :

*Contribution à la reconnaissance automatique de l'écriture
manuscrite arabe, application sur les montants
littéraux des chèques*

Présenté par :
BOUGAMOUZA Fateh

Dirigé par :
Pr. BENMOHAMMED Mohamed

Devant le jury composé de :

Dr. A. Chaoui	Université de Constantine	Président
Prof. M. Benmohammed	Université de Constantine	Rapporteur
Dr. S. Chikhi	Université de Constantine	Examineur
Dr. F. Hachouf	Université de Constantine	Examineur

Année Universitaire : 2008-2009

Remerciements

Je tiens à exprimer mes sincères remerciements à Mr. Med. Benmohammed, mon encadreur d'avoir accepté la tâche de rapporteur sur ce travail.

Je tiens à exprimer aussi mes sincères remerciements à :

Mr. A. Chaoui, maitre de conférences à l'université de Constantine, pour m'avoir fait l'honneur de présider le jury.

Mr. S. Chikhi, maitre de conférences à l'université de Constantine, pour accepter de consacrer du temps à mon travail et pour avoir participer à mon jury.

Mme. F. Hachouf, maitre de conférences à l'université de Constantine, pour accepter de consacrer du temps à mon travail et pour avoir participer à mon jury.

Mes derniers remerciements, à tous ceux qui, de prés ou de loin, par leur soutien et encouragement, m'ont aidé à réaliser ce travail.

ملخص

التعرف الآلي على النص المكتوب بأحرف متصلة لا يزال مشكلة قائمة في كل من شكله المخطوط وكذلك المطبوع؛ وذلك بسبب الصعوبات التي تواجه الباحثين والمطورين، مثل التغيير في الشكل والأسلوب، والميل في الكتابة. تتميز الكتابة العربية المخطوطة بكونها متصلة مما يجعلها صعبة المعالجة، وتبرز اختلافات كبيرة.

نظرا لطبيعة الكتابة العربية المطبوعة والمخطوطة في كونها متصلة ومرتبطة، العديد من نظم التعرف الآلي المطورة كان أساسها نماذج ماركوف المخفية؛ مختلف الأعمال المنجزة تستند إلى حد كبير على الخبرة المكتسبة في مجال التعرف على الكلام، حيث يتم عادة استخدام هذه النماذج بشكل متكرر. مساهمة هذه التقنية تكمن في حل المشاكل ذات الصلة بمعالجة اللغة العربية، مثل وجود المدود والصلوات والتجزئة.

في إطار العمل الذي قمنا به، والمتمثل في وضع نظام للتعرف الآلي على المبالغ الحرفية المخطوطة بحروف عربية وتطبيق هذا النظام على الشيكات من خلال التأليف بين مصنفات ماركوفية. اقترحنا أن نمثل الصورة ثنائية الأبعاد بإشارة ذات بعد واحد للاستفادة من التقدم المسجل في مجال معالجة الإشارات الصوتية أين نجد أن نماذج ماركوف يتم استخدامها بشكل متكرر. وبما أن الدراسات تظهر أن عمليات المعالجة الأولية المعروفة في معالجة الصور ليست كلها مناسبة للخط العربي اقترحنا تقنيات جديدة للمعالجة الأولية.

الكلمات المفتاحية: التعرف الآلي على الخط العربي، معالجة الإشارات، المعالجة الأولية للصور، نماذج ماركوف المخفية، التأليف بين المصنفات.

RESUMÉ

La reconnaissance des textes cursifs reste toujours un problème ouvert aussi bien dans sa forme imprimée que manuscrite. 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.

A cause de la nature cursive et connectée de l'écriture arabe imprimée et manuscrite, plusieurs systèmes basés sur les Modèles de Markov Cachés (MMCs) ont été développés pour la reconnaissance ; Les différents travaux réalisés reposent pour une grande part sur l'expérience accumulée dans le domaine de la reconnaissance de la parole où les MMCs sont fréquemment utilisés. L'apport de cette technique est dans la résolution des problèmes liés au traitement de l'Arabe tels que la présence de ligatures et la segmentation.

Dans le cadre de notre travail, qui consiste à développer un système de reconnaissance de l'écriture manuscrite arabe appliqué aux montants littéraux des chèques par combinaison de classifieurs markoviens. Nous avons proposé de représenter l'image binaire sous forme d'un signal unidimensionnel afin de tirer profit du progrès constaté dans le domaine de traitement du signal de la parole où les MMCs sont fréquemment utilisés. Et puisque la littérature montre que les opérations de prétraitements connues en traitement d'images, ne sont pas toutes appropriées à l'Arabe, des nouvelles techniques pour le prétraitement sont aussi proposées.

Les mots clés : Reconnaissance de l'écriture manuscrite arabe, traitement de signal, prétraitements des images, modèles de Markov cachés, combinaison de classifieurs.

Abstract

The recognition of the cursive texts remains always an open problem as well in its printed form as handwritten. This because of the difficulties with which the researchers and the developers are confronted, such as the variability of the form, the style, and the slant of the writing. The Arabic handwritten is naturally cursive, difficult to treat, and has a great variability.

Because of the cursive and connected nature of printed and handwritten Arabic script, several systems based on Hidden Markov models (HMM) were developed for the recognition; various realized works rest for a great part on the experience accumulated in the field of the voice recognition where HMMs are frequently used. The benefit of this technique is in the resolution of the problems involved in the treatment of the Arab such as the presence of ligatures and the segmentation.

In our framework, which consists in developing a system of recognition of the Arabic handwritten applied to the literal amounts of the checks, by combination of Markovian classifiers, We proposed to represent the binary image in the form of a one dimensional signal, in order to benefit from the noted progress in the field of treatment of the voice signal where HMMs are frequently used. And since the literature shows that the operations of preprocessing known in image processing are not appropriate to the Arabic, new techniques for preprocessing are also proposed.

Key words: The recognition of Arabic handwritten, signal processing, the image preprocessing, hidden Markov models, combination of classifiers.

SOMMAIRE

<i>Introduction générale</i>	01
<i>Chapitre1 La reconnaissance des formes</i>	03
1. Introduction	03
2. Définition	03
3. Les domaines d'application	04
Reconnaissance des signaux	04
Reconnaissance des formes dans les images numériques	04
4. Le processus de RdF	04
L'acquisition ou détection	06
Le prétraitement	06
Extraction des caractéristiques	06
Classification	07
4.4.1 L'apprentissage	07
4.4.1.1. L'apprentissage supervisé	07
4.4.1.2. L'apprentissage non supervisé (sans professeur)	08
4.4.2 Le test	08
4.4.3 Mesures de performances d'un classifieur	08
Post-traitement	09
5. Les approches de classifications	09
L'approche statistique	10
5.1.1. Exemples des méthodes statistiques	10
5.1.1.1.L'analyse discriminante linéaire	10
5.1.1.2. Les k plus proches voisins Kppv	11
5.1.1.3. Les réseaux de neurones	11
L'approche structurelle	12
L'approche hybride (ou l'approche statistico-structurelle)	13
Modèle de Markov caché	13
6. Combinaison de classifieurs	13
Topologie des ensembles de classifieurs	14
La fonction de décision	14
7. Conclusion	15
<i>Chapitre2 La reconnaissance automatique de l'écriture</i>	16
1. <i>Introduction</i>	16
2. Les systèmes de reconnaissance de l'écriture	16
2.1 Le mode d'acquisition de l'écriture	17
Systèmes de reconnaissance en ligne	17
Systèmes de reconnaissance hors ligne	17
2.2 Le type de l'écriture	18
2.2.1 Les systèmes de reconnaissance de l'écriture imprimée	18
2.2.2 Les systèmes de reconnaissance de l'écriture manuscrite	19
3. La reconnaissance automatique de l'écriture manuscrite (REM)	19
3.1 Les critères de complexité des systèmes de REM	19
3.1.1 La disposition spatiale et le style d'écriture	19
3.1.2 Le nombre de scripteurs	20

3.1.3	La taille du vocabulaire	20
3.2	Les approches de REM	21
3.2.1	Reconnaissance de caractères isolés	22
3.2.2	Reconnaissance de mots	22
3.2.2.1	L'approche globale	22
3.2.2.2	L'approche analytique	23
3.3	Étapes du processus de reconnaissance hors-ligne de mots manuscrits	23
3.3.1	Étape d'acquisition	24
3.3.2	Étape de pré-traitement	24
3.3.2.1	Binarisation	24
3.3.2.2	Redressement	25
3.3.2.3	Lissage	26
3.3.2.4	Filtrage	26
3.3.2.5	Squelettisation	26
3.3.2.6	Segmentation	26
3.3.3	Étape d'extraction de caractéristiques ou de primitives	27
3.3.3.1	Une première distinction	28
3.3.3.2	Une seconde distinction	28
3.3.4	Étape de classification	29
3.3.5	Étape de post-traitement	29
4.	Conclusion	29
	<i>Chapitre3 Le traitement automatique des montants littéraux arabes</i>	30
1.	Introduction	30
2.	Caractéristiques morphologiques de l'écriture arabe	31
	L'alphabet arabe	31
	Les caractéristiques de l'écriture arabe	32
3.	Quelques systèmes de traitement automatique des montants littéraux arabes	35
	Le système de Al-Ohli, 2002	35
	Le système de Al-Ma'adeed, 2004	37
	Le système de N. Farah 2005	39
	Le système de Souici 2005	42
4.	Conclusion:	45
	<i>Chapitre4 Les modèles de Markov cachés (MMCs)</i>	46
1.	Introduction	46
2.	Définition des MMCs	47
3.	Les éléments d'un MMC	48
4.	Extensions des MMCs	49
	Densités de probabilités	49
	Duré de séjour dans un état	49
	Ordre d'une chaîne	50
5.	Les types des MMCs	50
	MMC ergodique	50
	MMC gauche-droite	51
6.	Les problèmes fondamentaux pour les MMCs	51
	Problème d'évaluation	52
6.1.1	L'algorithme forward-backward	52

6.1.2	L'algorithme de Viterbi	54
	Problème de reconnaissance	54
	Problème d'apprentissage	55
	Algorithme de Baum-Welch	55
	Algorithme de Viterbi	56
7.	Les MMCs et la reconnaissance automatique de l'écriture	57
8.	Conclusion	60
	Chapitre 5 Contribution à La reconnaissance des montants littéraux arabes	61
1.	Introduction	61
2.	La base d'images	61
3.	Le système proposé	64
3.1	Acquisition	64
3.2	Prétraitements	64
3.2.1	Prétraitements classiques	65
3.2.1.1	Binarisation, filtrage	65
3.2.1.2	Suppression des espaces entourent le mot	65
3.2.2	Prétraitements proposés	65
3.2.2.1	Suppression des allongements horizontaux	65
3.2.2.2	Suppression des espaces verticaux	66
3.2.2.3	Suppression des espaces horizontaux (inter sous-mots).	66
3.3	Extraction des caractéristiques	67
3.3.1	Processus de conversion de l'image 2D en 1D	68
3.3.2	L'extraction des primitives pour le traitement de la parole	68
3.3.2.1	Les bancs de filtres	69
3.3.2.2	Le codage linéaire prédictif	69
3.3.2.3	Le cepstre	70
3.3.3	Extraction des caractéristiques dans notre système	71
3.4	Apprentissage	72
3.5	La reconnaissance	73
3.5.1	Stratégies de rejet	74
4.	Les algorithmes	76
5.	Résultats et interprétation	78
6.	Amélioration du système par combinaison des classifieurs	81
	Conclusion et Perspectives	83
1.	Conclusion générale	83
2.	Perspectives	84
	Références	85
	Annexe A	
	Annexe B	

La liste des figures

<i>Figure 1.1 : Schéma d'un système de reconnaissance de formes</i>	05
<i>Figure 2.1 : Écriture en ligne et hors ligne</i>	18
<i>Figure 2.2 : Classification des 5 styles d'écriture</i>	20
<i>Figure 2.3 Redressement de la ligne ; Le redressement des écritures penchées; Squelettisation</i>	25
<i>Figure 3.1 : Des variations dans l'écriture des montants littéraux</i>	31
<i>Figure 3.2 : Exemple de différentes formes de la boucle</i>	32
<i>Figure 3.3 : Le prétraitement</i>	36
<i>Figure 3.4 : Image d'un chèque de la base d'images CENPARMI</i>	37
<i>Figure 3.5 : Le montant littéral</i>	37
<i>Figure 3.6: Processus de reconnaissance en utilisant une combinaison de deux classifieurs</i>	38
<i>Figure 3.7: Des exemples des mots arabes appartenants aux montants littéraux</i>	39
<i>Figure 3.8: Le processus du système de reconnaissance</i>	40
<i>Figure 3.9: Exemple d'extraction de primitives globale</i>	43
<i>Figure 3.10 : Processus d'extraction des primitives</i>	43
<i>Figure 3.11 :L'intégration neuro-symbolique</i>	44
<i>Figure 4.1 : Un MMC ergodique</i>	50
<i>Figure 4.2 : Un MMC gauche-droite (a) Parallèle (b) Séquentiel</i>	51
<i>Figure 4.3: Illustration of the séquence of opérations required for the computation of the Forward variable $\alpha_{t+1}(j)$</i>	53
<i>Figure 4.4 : Illustration of the séquence of opérations required for the computation of the backward variable $\beta_t(i)$.</i>	54
<i>Figure 4.5 : Ecriture manuscrite et reconnaissance automatique</i>	57
<i>Figure 4.6 : Procédure d'utilisation d'un HMM pour la reconnaissance manuscrite</i>	58
<i>Figure 4.7 : HMM du mot « Eté »</i>	58

Figure 5.1 : Suppression des espaces entourent le mot «سبعمائة».	65
Figure 5.2 : (a) Elongations variables du mot «أربعة», (b) l'image du mot « اثنا » avant et après la suppression de l' allongement horizontal.	66
Figure 5.3 : l'image du mot « ثلاثمائة » avant et après la suppression des espaces verticaux	66
Figure 5.4 : l'image du mot « ثلاثمائة » avant et après la suppression des espaces horizontaux	67
Figure 5.5 : les probabilités lors des itérations d'apprentissage pour chaque mot	73
Figure 5.6 : Architecture du système proposé de reconnaissance de mot arabes	75
Figure 5.7 : échantillons de mots rejetés ou mal reconnus	80
Figure 5.8 : échantillon de mots mal reconnus	80
Figure 5.9 : échantillon de mot « و » très bien reconnu	81

La liste des tableaux

Tableau 3.1 : L'alphabet arabe dans ses différentes formes	34
Tableau 3.2: La distribution de sous-mots dans la base d'image	36
Tableau 3.3: Taux de reconnaissance du classifieur globale	39
Tableau 3.4: Taux de reconnaissance du classifieur local et du système total	39
Tableau 3.5: Taux de reconnaissance de chaque classifieur	41
Tableau 3.6 : Taux de reconnaissance de la combinaison	42
Tableau 3.7 : Taux de reconnaissance de chaque approche et d'hybridation	45
Tableau 4.1 : Comparaison entre DHMM et CHMM	49
Tableau 5.1 : Le vocabulaire des montants littéraux arabes	62
Tableau 5.2 Exemple des mots illustrant différentes difficultés relatives au manuscrit arabe dans notre base.	63
Tableau 5.3 : Les résultats avant et après l'utilisation des prétraitements proposés	78
Tableau 5.4 : Les taux de reconnaissance	79
Tableau 5.5 : les taux de reconnaissance par combinaison et sans prétraitements proposés.	82
Tableau 5.6 : les taux de reconnaissance par combinaison et avec prétraitements proposés.	82

La liste des algorithmes

<i>Algorithme 4.1 : L'algorithme Forward</i>	52
<i>Algorithme 4.2 : L'algorithme Backward</i>	53
<i>Algorithme 4.3 : L'algorithme de Viterbi</i>	55
<i>Algorithme 4.4 : L'algorithme Baum-Welch</i>	56
<i>Algorithme 5.1 : L'algorithme de suppression des allongements horizontaux</i>	76
<i>Algorithme 5.2 : L'algorithme de suppression des espaces verticaux</i>	76
<i>Algorithme 5.3 : L'algorithme de suppression des espaces horizontaux</i>	76
<i>Algorithme 5.4 : L'algorithme d'apprentissage du système proposé</i>	77
<i>Algorithme 5.5 : L'algorithme de test du système proposé</i>	77

La liste des équations

1.1 Taux de rejet	09
1.2 Taux de reconnaissance	09
1.3 Taux de substitution	09
1.4 Fonction discriminante linéaire	11
4.1 Chaîne de Markov d'ordre 1	47
4.2 Processus observable	47
4.3 La distribution de probabilité de transition	48
4.4 La distribution de probabilité de symboles d'observation	48
4.5 La distribution d'état initial	48
4.6 La contrainte 1 d'un MMC	48
4.7 La contrainte 2 d'un MMC	48
4.8 La contrainte 3 d'un MMC	48
4.9 MMC ergodique	51
4.10 MMC gauche-droite	51
4.11 Variable backward	53
5.1 Les valeurs dans le cas d'un pixel noir	68
5.2 Les valeurs dans le cas d'un pixel blanc	68
5.3 L'image sous forme d'un signal	68

5.4	Codage linéaire prédictif	69
5.5	Les coefficients du cepstre	70
5.6	Les fréquences de l'échelle Mel	70
5.7	Trois formes pour les dérivées des coefficients MFCC	71
5.8	les dérivées deuxièmes des coefficients MFCC	71

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 a de nombreuses applications, notamment en bureautique et tri automatique du courrier, pour l'enregistrement et la vérification des chèques bancaires, en publication assistée par ordinateur (PAO) pour faciliter la composition à partir d'une sélection de plusieurs documents et le traitement automatique de dossiers administratifs.

La reconnaissance des textes cursifs reste toujours un problème ouvert aussi bien dans sa forme imprimée que manuscrite. 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.

Les chèques bancaires et postaux font partie des documents les plus utilisés, et l'automatisation de son traitement devient un domaine de recherche important du point de vue économique et scientifique. Un chèque contient beaucoup d'informations qui sont souvent manuscrites, on cite parmi autres un montant numérique et un montant littéral. Ce dernier est le seul acceptable d'un point de vue légal. Cependant l'information redondante entre les deux montants (littéral et numérique) peut augmenter le taux de reconnaissance et par conséquent l'amélioration des performances du système.

A cause de la nature cursive et connectée de l'écriture arabe imprimée et manuscrite, plusieurs systèmes basés sur les Modèles de Markov Cachés (MMCs) ont été développés pour la reconnaissance ; Les différents travaux réalisés reposent pour une grande part sur l'expérience accumulée dans le domaine de la reconnaissance de la parole où les MMCs sont fréquemment

utilisés. L'apport de cette technique dans la résolution des problèmes liés au traitement de l'Arabe tels que la présence de ligatures et la segmentation.

Dans le cadre de notre travail, qui consiste à développer un système de reconnaissance de l'écriture manuscrite arabe appliqué aux montants littéraux des chèques. Nous avons proposé de représenter l'image binaire sous forme d'un signal unidimensionnel afin de tirer profit du progrès constaté dans le domaine de traitement du signal de la parole où les MMCs sont fréquemment utilisés. Et puisque la littérature montre que les opérations de prétraitements connues en traitement d'images, ne sont pas toutes appropriées à l'Arabe, des nouvelles techniques pour le prétraitement sont aussi proposées.

Notre mémoire se subdivise donc comme suit :

Dans le premier chapitre, nous présentons un aperçu sur la reconnaissance de formes et les étapes suivies pour la réalisation d'un système de reconnaissance en général.

Le deuxième chapitre est consacré à l'état de l'art dans le domaine de la reconnaissance automatique de l'écriture où nous décrivons les étapes nécessaires au développement d'un système de reconnaissance.

Le troisième chapitre se répartit en deux sections, la première étudie brièvement la langue arabe, et la seconde présente quelques systèmes de traitement automatique des montants littéraux arabes.

Le quatrième chapitre expose la théorie des modèles de Markov cachés, ses problèmes fondamentaux et leurs solutions. A la fin du chapitre, l'application de ces modèles pour la reconnaissance de l'écriture manuscrite est expliquée.

Le dernier chapitre décrit les étapes de la reconnaissance appliquées dans notre système et les expériences effectuées dans le cadre de ce travail et l'analyse des résultats obtenus.

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

Chapitre1

La reconnaissance des formes

Résumé: Dans ce chapitre, nous présenterons le domaine de la reconnaissance de formes (RdF). Tout d'abord nous donnerons sa définition, ces domaines d'application, et par la suite, nous exposerons le processus d'un système de RdF en détaillant ses étapes en cours du processus. Finalement, nous expliquerons les approches de classification (structurelle, statistique et hybride) avec quelques exemples et l'approche par combinaison des classifieurs.

1. Introduction

Chaque jour, et d'une manière intuitive, les hommes accomplissent des tâches de discernement des objets de forme connue. Par exemple, nous distinguons facilement, un triangle d'un cercle, un tigre d'un chat, le visage d'un enfant et celui d'un vieil. Pour ces tâches simples, l'homme est considéré comme un système relativement parfait comparé aux systèmes de reconnaissance artificielle.

Cependant, la reconnaissance devient très difficile et nécessite plus d'efforts et du temps dans d'autres cas comme : la lecture des montants des chèques postaux pour saisir en ordinateur, des adresses des lettres pour leur classification selon la ville, la recherche dans une base énorme d'empreintes digitales pour juger un suspect, la détermination d'identité d'un locuteur à partir de son voix. Pour cela, la reconnaissance automatique des formes revient non pas pour remplacer l'homme par une machine mais pour le soulager des tâches critiques et fastidieuses.

2. Définition

La reconnaissance de formes est un domaine pluridisciplinaire se situant au carrefour de plusieurs disciplines, on cite parmi autres : la linguistique, les statistiques, l'électronique, l'informatique, la théorie de communication, la biologie, l'optique,...etc [1]. Cette discipline consiste à reproduire les capacités perceptuelles de l'homme, en catégorisant un ensemble de données brutes selon le patron observé. La reconnaissance des formes ne se limite pas

seulement à la perception visuelle, mais inclut également tous les types de perception artificielle exigeant de reconnaître des patrons de données [2].

La reconnaissance doit répondre aux spécifications suivantes [3]:

- la reconnaissance doit être invariante par rotation et changement d'échelle des symboles et caractères (i.e. par similitude). En effet, de nombreuses chaînes de caractères et de symboles sont présentes en de multiples orientations et en plusieurs tailles.
- la reconnaissance doit être robuste vis à vis du placement des formes. Le recouvrement ou l'interconnexion de symboles doit pouvoir être analysé.

3. Les domaines d'application

Généralement, les objets à reconnaître sont de type signal ou image, ce qui implique deux grandes familles d'applications :

Reconnaissance des signaux

Où on essaye d'automatiser l'interprétation des signaux :

- signaux vocaux permettent d'automatiser le comportement auditif (l'écoute) d'un humain : reconnaissance de la parole, reconnaissance du locuteur...
- signaux biomédicaux : électrocardiogramme...
- surveillance d'instruments, diagnostics de panne...

Reconnaissance des formes dans les images numériques

Où on essaye d'automatiser le processus de la vision humaine :

- reconnaissance d'empreintes digitales,
- la reconnaissance de l'écriture,
- l'analyse de documents,
- imagerie satellitaires,
- analyse de scènes, interprétation d'images, « computer vision ».

4. Le processus de RdF

Le but du processus de RdF est la réduction de la quantité de données à manipuler, partant de l'information originale qui fait partie de l'espace d'observation ou monde réel

(souvent une image ou un signal) jusqu'à arriver à sa description symbolique, dans l'espace d'interprétation ou espace des catégories, en passant par l'espace de représentation ou espace des caractéristiques où en extraire les primitives pertinentes.

A partir de cette description, on déduit que dans le cas général, il existe deux étapes principales qui sont :

- l'étape d'extraction des caractéristiques : le passage de l'espace d'observation vers l'espace de représentation.

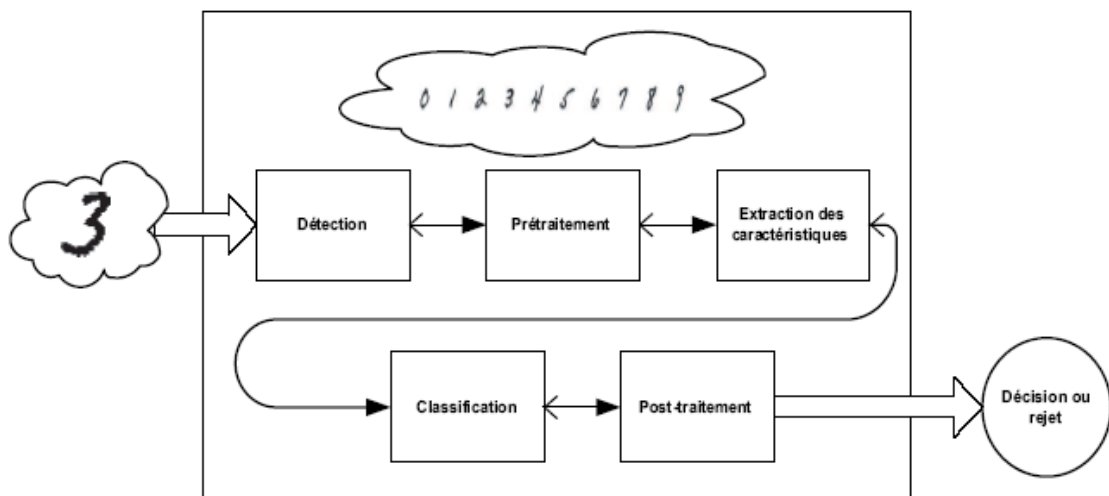


Figure 1.1 : Schéma d'un système de reconnaissance de formes : Précision sur le fonctionnement d'un système de reconnaissance de formes, ici divisé en 5 étapes. Le sens des flèches indique la principale direction du flux de données. Notons toutefois qu'une rétroaction est possible entre toutes ces étapes et que le schéma pourrait se complexifier avec des flèches partant de toutes les étapes pour revenir à n'importe quelle étape précédente [4].

- l'étape de classification : le passage de l'espace de représentation vers l'espace d'interprétation.

D'autres étapes qui sont complémentaires et parfois nécessaires tel que l'acquisition, le prétraitement et le post-traitement. Ces étapes sont décrites par [4] dans la figure 1.1.

L'acquisition ou la détection

Il faut d'abord acquérir l'information originale (la forme à reconnaître) à l'aide des capteurs physiques, et de la convertir à des grandeurs numériques, pour en mettre dans une chaîne de traitements informatisés.

Dans le cas où l'information est disponible sur un support souvent papier, les capteurs physiques sont des scanners ou des caméras numériques..., et dans le cas où la saisie effectuée en ligne le capteur est un écran tactile ou une tablette graphique munie d'un stylo électronique (stylet). Et pour capturer le signal vocal un microphone doit être utilisé, et la numérisation ensuite se fait à l'aide d'un convertisseur analogique-numérique.

Le prétraitement

L'objectif des prétraitements est de faciliter la caractérisation de la forme [5]. C'est l'étape de la préparation des données pour celle d'extraction des caractéristiques, elle agit sur deux axes, l'un est de réduire le bruit et les déformations ou de l'éliminer carrément s'il est possible ; et l'autre est de ne maintenir que les caractéristiques significatives de la forme et utiles dans les étapes suivantes.

Le traitement du signal et de l'image sont les principales sources pour les méthodes de prétraitement : filtrage (des bruits hautes fréquences, par exemple), amélioration des contrastes, extraction de contours ou de squelettes, modélisation du signal temporel, extraction des primitives, etc [6].

Extraction des caractéristiques

L'objectif de l'extraction des caractéristiques dans le domaine de la reconnaissance consiste à exprimer les primitives sous une forme numérique ou symbolique appelée codage. Selon le cas, les valeurs de ces primitives peuvent être réelles, entières ou binaires [5].

La difficulté ici est de trouver de bonnes caractéristiques. De « bonnes » caractéristiques permettent aux classifieurs de reconnaître facilement les différentes classes d'objets ; on dit alors qu'elles sont discriminantes. Elles doivent aussi être invariantes à certaines transformations (la lettre A appartient à la même classe quelle que soit sa taille) [4].

Classification

La classification est l'élaboration d'une règle de décision qui transforme les attributs caractérisant les formes en appartenance à une classe (passage de l'espace de codage vers l'espace de décision) [7].

Le type d'une méthode de classification se décline généralement en deux familles : le mode supervisé et le mode non supervisé. Si l'on dispose d'un ensemble de points étiquetés, on parlera de classification supervisée. Dans le cas contraire, on effectue une classification non supervisée appelée également classification automatique. La RdF, ainsi définie, est l'apprentissage ou la découverte de structures appelées classes dans un ensemble de données éventuellement perturbées [8].

Avant qu'un modèle de décision ne soit intégré dans un système de reconnaissance, il faut avoir procédé auparavant à deux étapes : l'étape d'apprentissage et l'étape de test [5].

L'apprentissage

L'étape d'apprentissage consiste à caractériser les classes de formes de manière à bien distinguer les familles homogènes de formes. C'est une étape clé dans le système de reconnaissance. On distingue deux types d'apprentissage : apprentissage supervisé et apprentissage non supervisé [7].

L'apprentissage supervisé

Dans le cas de l'apprentissage supervisé, un échantillon représentatif de l'ensemble des formes à reconnaître est fourni au module d'apprentissage. Chaque forme est étiquetée par un opérateur appelé professeur, cette étiquette permet d'indiquer au module d'apprentissage la classe dans laquelle le professeur souhaite que la forme soit rangée. Cette phase d'apprentissage consiste à analyser les ressemblances entre les éléments d'une même classe et les dissemblances entre les éléments de classes différentes pour en déduire la meilleure partition de l'espace des représentations. Les paramètres décrivant cette partition sont stockés dans une table d'apprentissage à laquelle le module de décision se référera ensuite pour classer les formes qui lui sont présentées [5].

L'apprentissage non supervisé

Dans le cas de l'apprentissage non supervisé, on fournit au système de reconnaissance un grand nombre de formes non étiquetées. L'étape de la classification va se charger d'identifier automatiquement les formes appartenant à une même classe.

Dans le domaine de la reconnaissance de l'écriture, ce sont les méthodes basées sur un apprentissage supervisé qui sont le plus souvent utilisées ; et plus particulièrement pour les caractères manuscrits isolés car les classes naturelles sont connues et en nombre limité [5].

Dans ce cas le nombre de classes peut être connu ou non, mais les classes des échantillons ne sont pas connues, ce qu'on sait c'est que l'ensemble des échantillons représente ce qu'on doit connaître. Une fois l'apprentissage effectué, le système doit être validé par une phase de test.

Le test

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 primitives ou le choix de la méthode d'apprentissage. En effet, il est difficile de trouver a priori les primitives pertinentes et la méthode d'apprentissage la plus adaptée au problème posé d'où l'utilité de procéder par itérations successives. Ces itérations consistent à extraire des primitives jugées utiles au problème de reconnaissance à résoudre et de tester la performance du système avec cet ensemble de primitives. Au fur et à mesure que les performances du système souhaitées ne sont pas atteintes alors il suffit de trouver à nouveau une nouvelle famille de primitives ou de combiner les primitives extraites avec de nouvelles primitives [5].

Le test d'une donnée se déroule en trois étapes [8]:

- acquisition des observations de l'objet à tester ;
- extraction des mêmes caractéristiques utilisées lors de l'apprentissage ;
- classement parmi les classes $C_1 ; \dots ; C_c$ par la règle de décision.

Mesures de performances d'un classifieur

Pour une entrée donnée, un classifieur peut générer les réponses suivantes [9]:

- **Un rejet** : pour indiquer que le classifieur n'a pas pu identifier cette entrée.

- **Une reconnaissance** : dans ce cas, il identifie bien l'entrée, et il lui attribue sa classe appropriée.
- **Une substitution** : le classifieur attribue une autre classe à l'entrée.

La performance d'un classifieur peut être mesurée en calculant les trois taux suivant :

$$\text{Taux de rejet} = \frac{\text{Nombre de formes rejetées}}{\text{Nombre total de formes}} \quad (1.1)$$

$$\text{Taux de reconnaissance} = \frac{\text{Nombre de formes reconnues}}{\text{Nombre total de formes}} \quad (1.2)$$

$$\text{Taux de substitution} = \frac{\text{Nombre de formes males reconnues}}{\text{Nombre total de formes}} \quad (1.3)$$

Post-traitement

L'étape ultime du processus de RdF, appelée post-traitement, regroupe toutes les actions à prendre lors de la classification ou non de l'objet à classer (fusion des sorties de plusieurs classifieurs, évaluation d'un seuil de confiance, décision de classer ou de rejeter un objet) [4], en utilisant d'autres informations de haut niveau (lexicales, syntaxiques ou sémantiques) pour sélectionner une solution.

5. Les approches de classifications

Les approches de classification sont généralement regroupées en deux grandes familles, l'une statistiques, et l'autre structurelle. Il est aussi possible de les combiner afin d'aboutir une approche hybride regroupant leurs avantages. Dans ce qui suit, on va donner une brève présentation de ces approches.

L'approche statistique

Une approche qui repose essentiellement sur des fondements mathématiques (probabilité & statistique). L'objet de cette approche est de décrire les formes à partir d'un modèle probabiliste simple à utiliser et de regrouper les formes dans des classes.

La théorie de la décision Bayésienne est la théorie centrale des méthodes statistiques, qui permet en fait de choisir l'hypothèse ayant la probabilité la plus élevée "plus forte probabilité d'appartenance à une classe".

La classification selon Bayes fait intervenir une distribution de probabilité par classe ; si la loi de probabilité $P_i(x)$ est connue pour chaque i , ainsi que la fréquence relative Q_i de chaque classe, alors les discriminateurs entre les classes peuvent être construits directement. Ils donnent lieu au classificateur de Bayes qui minimise le taux d'erreur de classement globale [2].

On distinguera d'une façon générale deux grands types de méthodes: les méthodes dites *non paramétriques*, où l'on cherche à définir les frontières des classes dans l'espace de représentation, de façon à pouvoir classer le point inconnu par une série de tests simples; les méthodes dites *paramétriques* ou *bayésiennes*, où l'on se donne un modèle de la distribution de chaque classe (en général gaussien), et où l'on cherche la classe à laquelle le point a la probabilité la plus grande d'appartenir. Dans le premier cas, une tactique simple est de définir des hyperplans qui séparent au mieux les classes d'apprentissage. La décision se réduit alors à une série de produits scalaires. Ces méthodes sont dites de *séparation linéaire*. Une autre approche non paramétrique très utilisée est celle de la décision par plus proche voisin: on attribue au point inconnu la classe de son plus proche voisin de l'ensemble d'apprentissage [10].

Les méthodes statistiques sont relativement simples et peu coûteuses en temps de calcul surtout pour l'approche paramétrique [11], et moyennement sensibles au bruit.

5.1.1. Exemples des méthodes statistiques

5.1.1.1. L'analyse discriminante linéaire

L'analyse discriminante linéaire permet de prendre en compte une telle information. Cette technique statistique consiste à chercher les composantes ou directions efficaces,

permettant de réaliser la discrimination des données, en fonction des modalités d'une variable qualitative.

Ces composantes permettent alors d'effectuer une projection des données dans un nouvel espace (transformation linéaire de l'original) dans lequel les échantillons sont regroupés en fonction des modalités de la variable qualitative. Son application, en utilisant l'information de classe comme variable qualitative, conduira à la construction d'un espace de représentation plus discriminant [12].

5.1.1.2. Les k plus proches voisins Kppv

La méthode de Kppv ou en anglais Knn(K-Nearest Neighbors) est liée à la notion de proximité, avec les propriétés suivantes :

- elle ne fait pas d'hypothèse sur la forme des classes.
- possède des propriétés de convergences très intéressantes.

L'algorithme Kppv affecte une forme inconnue à la classe de son plus proche voisin en la comparant aux formes stockées dans une classe de référence nommée prototype. Il renvoie les K formes les plus proches de la forme à reconnaître suivant un critère de similarité. Une stratégie de décision permet d'affecter des valeurs de confiance à chacune des classes en compétition et d'attribuer la classe la plus vraisemblable (au sens de la métrique choisie) à la forme inconnue [13].

5.1.1.3. Les réseaux de neurones

Sont des méthodes connexionnistes, parfois considérées comme des méthodes statistiques. Elles permettent d'étendre les techniques de séparation linéaire (le cas du perceptron) à des séparations non-linéaires [6].

Les classifieurs de type réseaux de neurones proviennent d'une modélisation des neurones biologiques (McCulloch & Pitts, 1943). Ils reposent sur une modélisation discriminante. Un neurone permet de définir une fonction discriminante linéaire g dans l'espace de représentation E des entrées. Cette fonction réalise une combinaison linéaire du vecteur de caractéristiques de l'entrée e

$$g(e) = w^t e + w_0, \quad (1.4)$$

Où w est un vecteur de poids de la combinaison linéaire et w_0 est le biais. Ainsi $g(e) = 0$ définit un hyperplan permettant de séparer E en deux régions de décision. En déterminant les bonnes valeurs de w et w_0 , celles-ci peuvent être associées à deux classes pour faire leur discrimination. Pour des problèmes à n classes, n fonctions discriminantes sont établies. L'extension à des réseaux de type perceptron multi-couches (PMC) permet d'obtenir des frontières de décision de complexité quelconques.

L'apprentissage des poids est généralement fait par des méthodes comme la rétro-propagation du gradient de l'erreur.

Les PMC ont été très largement utilisés en classification et en reconnaissance de formes notamment pour leurs bonnes performances et leur simplicité [14].

L'approche structurelle

A l'inverse de l'approche statistique basée sur des métriques quantitatives où nous oublions les natures des objets, ces derniers sont représentés dans l'approche structurelle (syntaxique) de manière plus proche de la réalité, en se basant sur des mesures qualitatives extraits de la structure de ces objets.

La notion de structure, quoique sujette à de nombreuses définitions, fait toutefois apparaître l'existence d'une décomposition du tout (la structure) en parties et de relations entre ces parties. Les méthodes structurelles de reconnaissance de formes adoptent des représentations faisant explicitement apparaître les relations entre éléments de la forme. Elles sont évidemment plus appropriées pour traiter des structures visuelles où les composantes sont liées par des relations spatiales [6].

Dans l'approche structurelle, une classe est décrite par une grammaire. Une grammaire est composée d'un ensemble de règles syntaxiques qui déterminent l'ensemble des formes admissibles dans cette classe et présentant en principe des caractéristiques structurelles communes. Cet ensemble de formes est appelé langage engendré par la grammaire [1].

Les caractéristiques d'entrée des classificateurs basés sur l'approche structurelle appelées primitives ne sont pas constituées de vecteurs de valeurs réelles mais plutôt d'une liste d'attributs semi-objectifs (grand, ouvert, bleu pâle, etc.), de valeurs booléennes ou d'une

décomposition de l'objet en parties élémentaires idéalisées (ligne verticale, cercle, phonème, etc.) [4].

Font partie de cette catégorie : les arbres de décision, les systèmes experts et les programmes d'analyse syntaxique [4].

L'approche hybride (ou l'approche statistico-structurelle)

Dans [11,15]; les auteurs utilisent une collaboration entre approche statistique et structurelle (une approche statistico-structurelle). Ces systèmes utilisent deux niveaux d'interprétation. Le premier extrait puis reconnaît les primitives des formes grâce à un classifieur statistique. Le second exploite les caractéristiques topologiques des formes pour reconstruire des graphes (d'occlusions ou squelette) pour chacun des formes, puis reconnaît les formes à l'aide d'un classifieur structurel.

Modèle de Markov caché

Les modèles de Markov connaissent actuellement un essor important en reconnaissance des formes grâce à leur capacité d'intégration du contexte et d'absorption du bruit. Dans ces modèles, les formes sont décrites par une séquence de primitives qui seront observées dans les états du modèle. La probabilité d'émission de la forme par le modèle est calculée en maximisant, sur l'ensemble des chemins d'états, la probabilité d'observation des segments pondérée par les probabilités de transitions entre états. Ce calcul se fait généralement par maximum de vraisemblance. Le calcul de la vraisemblance de la forme par rapport au modèle intervient dans la règle de Bayes qui inclue la probabilité a priori du modèle [16].

6. Combinaison de classifieurs

Un système multi-classifieurs (Multiple Classifier System : MCS) est constitué d'un ensemble de différents classifieurs et d'une fonction de décision pour combiner leurs sorties. La description d'un MCS suit les deux phases suivantes [9]:

- Générer un ensemble de classifieurs complémentaires qui peuvent être combinés pour arriver à une solution optimale.
- Définir la fonction de combinaison pour donner une décision finale.

Topologie des ensembles de classifieurs

Les principales topologies utilisées sont [4]:

- **Conditionnelle** : Dans ce type d'organisation, un premier système de classification prend la décision de classer ou de rejeter un objet. Si ce dernier est rejeté, un second système prend la relève et ainsi de suite. En général le premier étage est constitué d'un classifieur efficace en terme de ressources (mémoire ou temps de calcul). Son rôle est de classer les objets les plus faciles à reconnaître et de laisser aux autres étages, plus complexes, les objets difficiles. C'est une façon de minimiser l'utilisation des ressources ou de réduire le temps moyen de reconnaissance d'un système.
- **Sérielle** : Il s'agit ici de construire un système où les classifieurs sont utilisés les uns à la suite des autres, chacun réduisant l'ensemble de classes possibles pour un objet à classer. Ainsi, un classifieur étant indécis et hésitant entre N classes passera l'information au classifieur qui le suit. Ce dernier produira un ensemble de M $M < N$ classes possibles et ainsi de suite jusqu'au dernier classifieur qui devra n'en retenir qu'une seule. C'est une façon de réduire un problème complexe en une série de problèmes plus simples.
- **Parallèle** : L'organisation parallèle des experts d'un ensemble de classifieurs implique que chaque classifieur individuel produise une sortie simultanée. L'ensemble de ces sorties est ensuite fusionné de façon à produire une décision unique. Si cette dernière organisation est la topologie la plus courante, c'est aussi celle qui a le plus grand coût en ressources.

La fonction de décision

Il existe plusieurs méthodes de combinaison des classifieurs, dont quelques-unes sont présentées ci-dessous [4]:

- **Vote à majorité simple** : C'est la fonction de fusion la plus simple et probablement la plus utilisée. La sortie de chaque expert est considérée comme étant un vote pour une classe. Le nombre de votes pour chacune des classes est compté et l'ensemble choisit la classe en ayant remporté le plus.
- **Sommation** : La sortie de chaque classifieur doit être un vecteur de taux de confiance, ou une probabilité a posteriori (à chaque classe correspondant un taux de confiance). La fusion est obtenue en sommant les vecteurs de sortie de tous les experts, puis en sélectionnant la classe ayant la valeur la plus élevée.

- **Règle du produit** : Comme la règle de sommation, l'utilisation de cette règle nécessite des sorties de classifieurs représentant un vecteur de taux de confiance. La fusion est obtenue en faisant le produit de ces taux pour chacune des classes, puis en sélectionnant la classe ayant la valeur la plus élevée. Cette méthode a le désavantage de produire des valeurs nulles dès qu'un classifieur produit une probabilité a posteriori nulle pour une classe donnée.

7. Conclusion

Nous avons présentés brièvement le domaine de RdF, leurs centres d'intérêts et domaines d'application, son processus, ainsi les approches utilisées, notamment dans l'étape de classification.

Au cours des quatre dernières décennies, de nombreuses méthodes pour l'analyse et la classification de formes ont été aboutis. Cependant, le problème de l'intelligence artificielle en général et de la RdF en particulier est si vaste qu'il est nécessaire de l'aborder par sous-problèmes. Parmi ceux-ci, la reconnaissance de l'écriture qui est un sous domaine de la RdF dans lequel les recherches sont particulièrement actives.

Chapitre2

La reconnaissance automatique de l'écriture

Résumé : *Nous développerons ce chapitre selon le schéma suivant. Dans un premier temps nous présenterons la classification des systèmes de reconnaissance de l'écriture suivant le mode d'acquisition partant par le mode en ligne. Ensuite nous exposerons le mode hors-ligne dans l'écriture imprimée et celle manuscrite. Finalement, nous détaillerons les différentes étapes d'un système de reconnaissance de l'écriture en mode hors-ligne.*

1. Introduction

L'écriture est un moyen de communication entre individus. L'homme rêve aussi de communiquer avec la machine de la même façon qui se semble plus simple et plus facile pour l'accélération des opérations d'accès, d'échange et de traitement des informations. D'où l'émergence de la reconnaissance automatique de l'écriture (RAE).

La RAE s'affirme de jour en jour comme une réelle technique de codage du support écrit et semble combler petit à petit son retard par rapport à d'autres disciplines voisines, comme la reconnaissance automatique de la parole. Concernant le script lui-même, la technique semble réduire largement l'écart qui existait entre les performances obtenues sur l'imprimé et celles obtenues sur le manuscrit [17].

La reconnaissance de l'écriture manuscrite a connues ces dernières années de grands progrès, et les succès des travaux de recherches ont donné lieu à de nombreuses applications industrielles, notamment dans le domaine de la lecture automatique de formulaires, de chèques ou d'adresses postales, ainsi que les applications de reconnaissance de l'écriture dites « en ligne » à travers les PDA, tablet-PC ou stylo caméra [18].

2. Les systèmes de reconnaissance de l'écriture

Les systèmes de reconnaissance de l'écriture peuvent être classifiés selon deux critères :

2.1 Le mode d'acquisition de l'écriture

Selon ce critère, les systèmes de reconnaissance automatique de l'écriture sont regroupés en deux familles :

Systèmes de reconnaissance en ligne

Dans ce type de systèmes, la reconnaissance est effectuée en temps réel, c'est-à-dire elle est effectuée au fur et à mesure que le caractère est tracé, ce qui permet d'obtenir une large marge de correction et modification selon la réponse donnée à la phase de reconnaissance chevauchée à la phase d'acquisition.

Ce mode d'acquisition est réservé généralement à l'écriture manuscrite. C'est une approche « signal » où la reconnaissance est effectuée sur des données à une dimension. L'écriture est représentée comme un ensemble de points dont les coordonnées sont fonction du temps [19]. Où on s'intéresse aux méthodes et techniques de traitement du message tel qu'il est écrit, en prenant en compte les informations relatives au mécanisme d'écriture telles la position des points, la vitesse et l'accélération qui sont des fonctions du temps [20, 17].

Les moyens de saisie en ligne sont nombreux où la tablette graphique avec un stylo électronique et l'écran tactile sont couramment utilisés.

Parmi les plus récentes plate-formes disposant d'un système de reconnaissance de l'écriture, nous trouvons le Palm et l'agenda électronique. Ces deux appareils regroupent une tablette à numériser et un programme procédant à la reconnaissance de l'écriture. De ce fait, son utilisation est plus attrayante puisqu'elle épargne à l'utilisateur le besoin de « scanner » a priori son écriture. Ceci a remis la reconnaissance en ligne au centre d'intenses efforts de développement au sein de la communauté de l'écrit et du document [21].

Systèmes de reconnaissance hors ligne

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, ayant perdu toute information temporelle sur l'ordre des points. De plus, ce mode introduit une difficulté supplémentaire relative à la variabilité du

tracé en épaisseur et en connectivité, nécessitant l'application de techniques de prétraitement [17].

Les domaines d'application les plus typiques sont principalement associés au traitement automatique des adresses postales, du montant des chèques, des formulaires, des feuilles de soins...

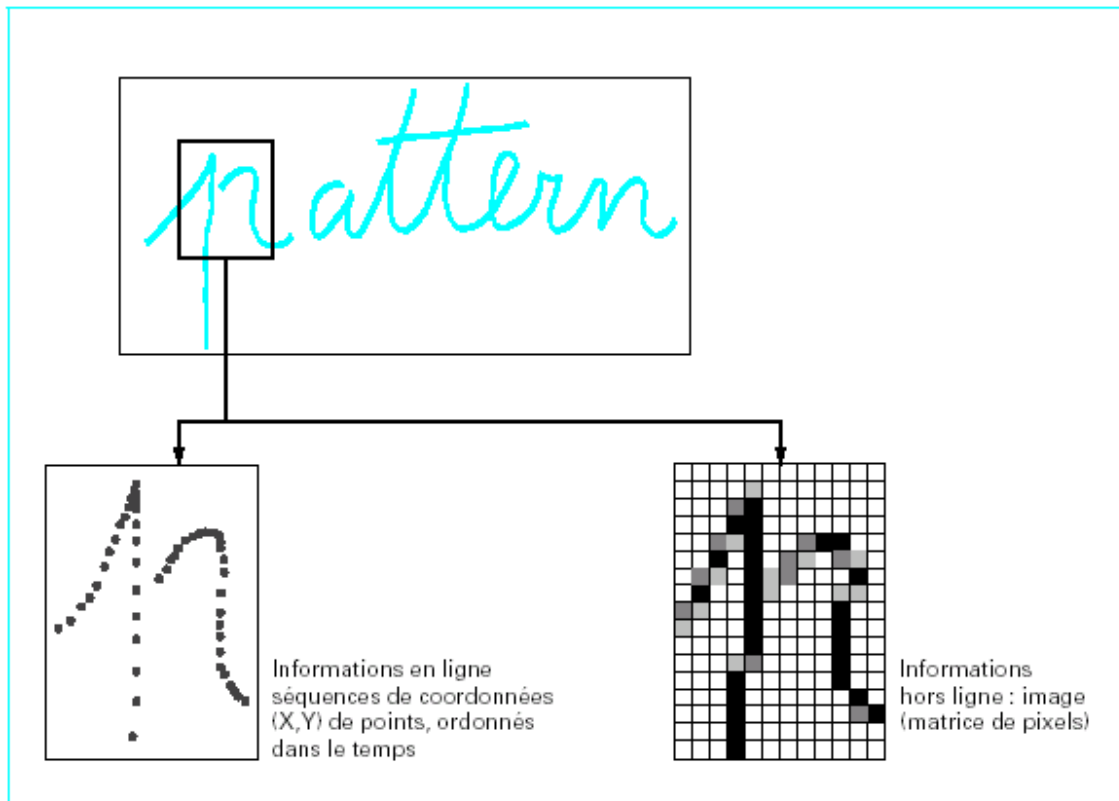


Figure 2.1 : Écriture en ligne et hors ligne [22]

2.2 Le type de l'écriture

Les systèmes de reconnaissance de l'écriture imprimée

Dans le document présenté au système de reconnaissance de l'écriture imprimée, l'écriture est caractérisée par le nombre de fontes (mono, multi ou omni-fonte). Le système de reconnaissance dans ce cas est généralement appelé OCR (pour Optical Character Recognition).

Un système OCR est dit mono-fonte s'il ne traite qu'une seule fonte à la fois, c'est-à-dire qu'il ne connaît que le graphisme d'une fonte unique. Ce cas est simple particulièrement

au niveau de l'apprentissage car l'alphabet représenté est réduit. Un système OCR est dit multi-fonte s'il est capable de reconnaître divers types de fontes parmi un ensemble de fontes préalablement apprises. Dans ce cas, le traitement doit réduire les écarts entre même caractère (taille, épaisseur et inclinaison). Un système OCR omni-fonte est capable de reconnaître toute fonte, généralement sans apprentissage, en se basant sur les règles topologiques et morphologiques de l'écriture [20].

Les systèmes de reconnaissance de l'écriture manuscrite

Une des spécificités de la reconnaissance de l'écriture manuscrite est la grande variabilité intra-classe associée aux données. Cette dernière est à imputer principalement aux scripteurs, qui possèdent chacun un style propre [12].

Dans le cadre de notre projet, nous nous sommes intéressés plus particulièrement à la reconnaissance de l'écriture manuscrite, qui est souvent considérée plus difficile et complexe que celle de l'écriture imprimée. Le reste de ce chapitre sera donc consacré au ce type d'écriture.

3. La reconnaissance automatique de l'écriture manuscrite (REM)

3.1 Les critères de complexité des systèmes de REM

D'une manière générale, la complexité de la reconnaissance d'information manuscrite dépend de plusieurs critères [17, 18, 22]:

3.1.1 La disposition spatiale et le style d'écriture

Selon la classification de Tappert [23]; la présentation du texte peut subir deux types de contraintes : *externes* conduisant à une écriture *pré-casée* pour lequel le scripteur doit s'efforcer d'écrire dans des cases prédéfinies (cas des formulaires, code postal d'une adresse), *zonée* (ou dans *des cadres* : montants de chèques), *guidée* (ou sur des lignes supports : lignes d'un bloc adresse, montant littéral d'un chèque), ou *générale* (ou bien *non conditionnée* : documents libres, par exemple des notes manuscrites) ; et *internes* provenant des habitudes propres à chaque scripteur et la difficulté à reconnaître l'écriture augmente avec les 5 styles d'écriture suivants(voir figure 2.2): écriture scripte pré-casée, écriture scripte avec caractères espacés, écriture scripte libre, écriture cursive, écriture mixte cursive et scripte. Notons

également la présence d'alphabets non latins plus délicats à reconnaître car possédant un grand nombre de symboles : caractères chinois, japonais, coréen, arabes,... etc.

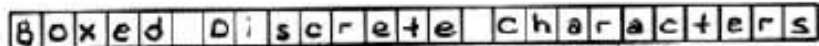
1.  1. Boxed Discrete Characters
2. Spaced Discrete Characters
3. Run on Discretely written Characters
4. Pure Cursive Script Writing
5. Mixed Cursive and Discrete

Figure 2.2 : Classification des 5 styles d'écriture du plus facile (1) au plus difficile (5) [23].

3.1.2 Le nombre de scripteurs

La réduction du nombre de scripteurs potentiels permet éventuellement de réduire la variabilité et d'apprendre les différents styles d'écriture. La difficulté de reconnaissance croît avec ce nombre et elle s'accroît en contexte omni-scripteur en raison des styles d'écriture très différents de chacun.

- **Mono-scripteur** : un seul scripteur peut utiliser le système de reconnaissance après apprentissage de son écriture ;
- **Multi-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.

3.1.3 La taille du vocabulaire

Les systèmes de reconnaissance de textes sont souvent basés sur un lexique qui facilite grandement la lecture, surtout si celui-ci possède un faible nombre de mots. Notons que dans le cas de la reconnaissance de séquences numériques, la présence d'un lexique est plus rare (cas de la reconnaissance de codes postaux).

Selon ce critère on distingue les trois types de systèmes suivants :

- **système de reconnaissance petit vocabulaire** : où le nombre de mots (ou de symboles) à reconnaître constitue un lexique de taille réduite (inférieure à 100 mots), (cas des montants littéraux de chèques qui contiennent une trentaine de mots).

- **système de reconnaissance à vocabulaire moyen:** contenant quelques centaines de mots.
- **système de reconnaissance grand vocabulaire** (étendu): plusieurs milliers ou dizaines de milliers de mots.

Il est évident que dans le premier cas, la complexité est moindre, car la réduction du nombre limite l'encombrement en mémoire et favorise l'utilisation de méthodes de reconnaissance directes et donc rapides, par balayage systématique de l'ensemble des mots du lexique.

3.2 Les approches de REM

L'objectif de la reconnaissance de l'écriture manuscrite est de développer un système qui se rapproche le plus de l'être humain dans sa capacité de lire [5]. Cependant, le problème de la reconnaissance de l'écriture en général et de l'écriture manuscrite en particulier est si vaste qu'il est nécessaire de l'aborder par des sous-problèmes.

Cette reconnaissance regroupe deux thèmes : la reconnaissance de caractères manuscrits isolés (numériques ou alphanumériques) et la reconnaissance de mots.

Il existe une autre distinction à faire entre le texte (caractères) continu et les caractères isolés. Ces derniers sont présents spécialement dans les formulaires où la zone d'écriture est restreinte. Par contre, le texte continu nécessite une phase de segmentation préliminaire à la classification [21].

La reconnaissance Hors-Ligne de caractères manuscrits isolés (numériques ou alphanumériques) est un domaine de recherche exploré depuis déjà plusieurs décennies, et débouche sur un nombre important d'applications industrielles. Ces applications visent la lecture automatique de documents manuscrits du type chèques bancaires ou courrier postal. En outre, bon nombre de traitements liés aux particularités de l'écriture manuscrite *non contrainte* (la reconnaissance de mots) sont encore aujourd'hui des problèmes ouverts dans le domaine de la reconnaissance de l'écrit [22,5].

3.2.1 Reconnaissance de caractères isolés

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) [17].

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 [5].

Deux types de reconnaissance de caractères numériques manuscrits [24]:

- la reconnaissance des chiffres isolés : peut être considéré comme un problème de classification des images des caractères isolés dans un ensemble d'alphabet donné.
- reconnaissance des chaînes numériques manuscrites : il est à noter que ce type de 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.

3.2.2 Reconnaissance de mots

Deux approches s'opposent en reconnaissance des mots : globale et analytique [17, 22]:

3.2.2.1 L'approche globale

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 [25].

L'approche globale est 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).

3.2.2.2 L'approche analytique

Contrairement à l'approche globale, l'approche analytique cherche à identifier les caractères ou 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*". 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 [17].

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 [26].

3.3 Etapes du processus de reconnaissance hors-ligne de mots manuscrits

Comme la reconnaissance de l'écriture est une branche de la reconnaissance des formes, elle suit le même processus d'un système de RdF présenté dans le chapitre précédent avec des particularités liées aux caractéristiques de l'écriture.

3.3.1 Etape d'acquisition

Les capteurs physiques CCD (Charge Couple Device) par exemple Scanner, caméras et appareils photos sont les appareils de numérisation les plus utilisés pour les systèmes de reconnaissance de l'écriture hors-ligne.

Un appareil de numérisation est défini par:

- nombre de signaux pouvant enregistrer dans un espace x, y (nombre de pixels qu'ils peuvent enregistrer).
- dynamique du signal : nombre de niveaux de gris ou de couleurs (de 256 couleurs à 16,7 millions, et plus).

Chaque appareil possède une résolution de prédilection pour laquelle il a été conçu, cette résolution peut être fixe ou variable (utiliser des fonctionnalités de type zoom... pour la modifier) ; les systèmes de reconnaissance travaillent généralement à des résolutions de 200dpi à 300dpi, résolutions suffisantes pour garder les principaux traits d'écriture avec un minimum de dégradation possible.

3.3.2 Etape de pré-traitement

Dans le but d'automatiser la reconnaissance de l'écriture, il faut préparer le travail de la machine. Beaucoup d'efforts ont été consacrés aux étapes préliminaires, qui sont nécessaires à la reconnaissance, parmi lesquelles on peut noter : la binarisation, le redressement et la squelettisation ...etc.

3.3.2.1 Binarisation

La binarisation est une étape importante pour la reconnaissance, elle consiste à détacher le texte du fond qui n'est pas utile à la reconnaissance, ce qui permet de diminuer la quantité de données à traiter et de réduire l'espace mémoire et le temps de calcul.

La binarisation permet de passer d'une image de niveaux de gris à une image binaire composée de 2 valeurs 0 et 1, plus simple à traiter. En général, on utilise un seuil de binarisation approprié qui traduit la limite des contrastes fort et faible dans l'image. Mais pour des images peu contrastées ou à contraste variable (i.e. la distribution de niveaux de gris n'est pas clairement bimodale), il est difficile de fixer ce seuil à une valeur précise [17].

3.3.2.2 Redressement

Le redressement est une technique de correction de l'inclinaison qui a été développée pour répondre essentiellement à deux besoins :

- corriger le défaut de positionnement du papier dans le scanner.
- corriger l'inclinaison interne de certaines parties du document qui peuvent apparaître (fantaisie de présentation, etc.)

En effet, la plupart des algorithmes de reconnaissance ne fonctionnent bien que si le document est bien redressé.

Le redressement se fait selon deux niveaux [17]:

- **le redressement de la ligne de base** : l'idée est de rendre horizontaux les mots à l'aide d'une transformation géométrique de type rotation isométrique des points de l'image (voir Figure 2.3 a).

- **le 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 (voir Figure 2.3 b).

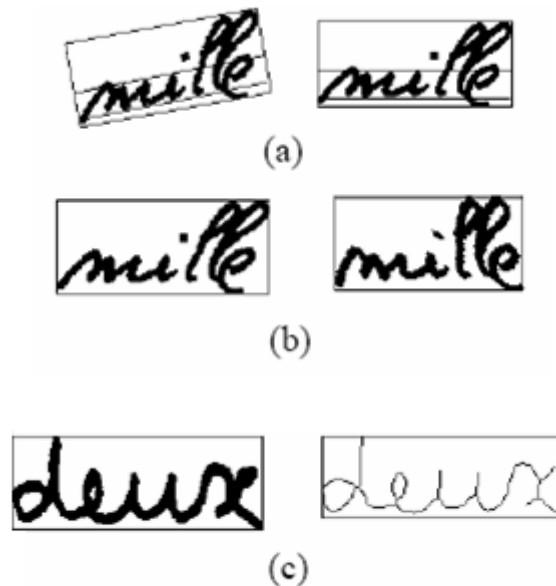


Figure 2.3 a-redressement de la ligne ;b- Le redressement des écritures penchées
c- Squelettisation [17]

3.3.2.3 Lissage

L'image des caractères peut être entachées 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. Les techniques de lissage permettent de résoudre ces problèmes par des opérations locales qu'on appelle opérations de bouchage et de nettoyage [27].

L'opération de nettoyage permet de supprimer les petites tâches et les excroissances de la forme. Pour le bouchage il s'agit d'égaliser les contours et de boucher les trous internes à la forme du caractère en lui ajoutant des points noirs. Plusieurs autres techniques similaires sont utilisées dont la méthode statistique, une méthode basée sur la morphologie mathématique [13].

3.3.2.4 Filtrage

Il existe plusieurs types de filtrage, on cite les suivants :

- filtrage linéaire : sert au rehaussement, la détection de contour, lissage passe-bas ...
- filtrage morphologique : l'élimination du bruit, la restauration des composantes connexes, et l'extraction du contour dans une image binaire peuvent se faire par des opérateurs morphologiques (éléments structurants).

3.3.2.5 Squelettisation

La squelettisation sert à obtenir une épaisseur égale à 1 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 (voir Figure 2.c) ; et donc en simplifier la représentation [17].

3.3.2.6 Segmentation

La segmentation est une opération très importante pour le problème de reconnaissance des caractères [28], des pseudo-mots et des mots.

La séparation des lignes, des mots, des pseudo-mots, des caractères constituent des opérations délicates et coûteuses, tant les écritures sont variées, les lignes sont chevauchés et

les caractères souvent liées (cas de l'arabe : l'écriture est semi-cursive) les unes aux autres. Une fois ces opérations faites, il reste à reconnaître chaque mot ou caractère, ce qui est également difficile.

La segmentation consiste de générer un découpage de l'image en des éléments susceptibles d'être reconnus, en décortiquant par exemple le caractère en sous caractères simples à identifier [28].

Selon l'unité de reconnaissance, Il y a deux approches principales, l'une globale dite aussi *holistique* où l'unité de base est le mot; et l'autre analytique ou dite *locale* utilise les graphèmes ou les caractères comme unités de base pour la reconnaissance. Mais, il faut d'abord segmenter la page et séparer les lignes.

- **segmentation de la page** : Consiste à déterminer les zones d'informations dans chaque page. Cette segmentation permet d'étiqueter la nature du média représenté dans chaque zone (texte, graphique, photographie,...etc).

- **segmentation d'un bloc de texte en lignes** : Le but de cette étape est la séparation des différentes lignes du texte afin de faciliter les autres niveaux de segmentation. En s'appuyant souvent sur les histogrammes de projection horizontale ou sur d'autres méthodes spécialisées comme la décomposition de l'image en composantes connexes. Cependant, le chevauchement et l'interférence des ascendants et des descendants des lignes voisines rendent difficile l'opération de segmentation.

- **segmentation des lignes en mots ou en sous-mots (pseudo-mots)** : En s'appuyant sur la détection des espaces entre les mots et les pseudo-mots en utilisant les projections verticales, l'objectif de cette étape étant la séparation des mots ou pseudo-mots selon l'approche suivie (l'approche globale ou l'approche pseudo-analytique).

- **segmentation des mots en caractères ou graphème** : La segmentation en caractères (ou en parties de caractères), constitue le problème le plus ardu lié à la reconnaissance de l'écriture. Cette tâche est opérée sur le squelette de la chaîne considérée [29].

3.3.3 Etape d'extraction de caractéristiques ou de primitives

L'objectif commun de toutes les primitives est de caractériser au mieux la forme des caractères, afin de pouvoir distinguer si deux images appartiennent à deux classes différentes

ou à la même classe, c'est-à-dire qu'elles doivent diminuer la variabilité intra-classe et augmenter la variabilité inter-classe. Suivant les applications et les techniques utilisées pour le système de reconnaissance, les primitives extraites peuvent être très différentes [30]. Dans la littérature, les primitives sont classifiées de plusieurs façons.

3.3.3.1 Une première distinction

La première distinction peut être effectuée entre les primitives globales et les primitives locales.

- **les primitives globales** : cherchent à représenter au mieux la forme générale d'un caractère et sont donc calculées sur des images relativement grandes (ex : transformée de Fourier et transformée de Hough).
- **les primitives locales** : sont calculées lors d'un parcours des pixels de l'image avec un pas d'analyse qui dépend de la modélisation, du type de primitive et de la taille de l'image.

3.3.3.2 Une seconde distinction

La seconde distinction est effectuée entre les primitives topologiques, structurelles ou statistiques [12,29]:

- **les primitives topologiques ou métriques** : elles consistent à compter dans une forme le nombre de trous, évaluer les concavités, mesurer des pentes et autres paramètres de courbures et évaluer des orientations, mesurer la longueur et l'épaisseur des traits, détecter les croisements et les jonctions des traits, mesurer les surfaces et les périmètres, ...
- **les primitives structurelles** : elles ressemblent beaucoup aux primitives topologiques. La différence est qu'elles sont généralement extraites non pas de l'image brute, mais à partir du squelette ou du contour de la forme. Ainsi, on ne parle plus de trous, mais de boucles ou de cycles dans une représentation filiforme du caractère.
- **les primitives statistiques** : l'histogramme, qui représente le nombre de pixels sur chaque ligne ou colonne de l'image, en est un exemple classique et simple à calculer. On peut citer également l'approche basée sur un moyennage des pixels situés à l'intérieur d'un masque rectangulaire : on construit une matrice de masques recouvrant la totalité de la forme qui permet une représentation statistique des valeurs correspondant à chaque masque.

3.3.4 Etape de classification

La classification est l'élaboration d'une règle de décision qui transforme les attributs caractérisant les formes en appartenance à une classe (passage de l'espace de codage vers l'espace de décision) [7, 5]. La classification se fait selon le même principe présenté dans le chapitre 1.

3.3.5 Etape de post-traitement

Le post-traitement correspond au traitement ultérieur de la classification ; il est effectué quand le processus de reconnaissance aboutit à la génération d'une liste de lettres ou de mots possibles. Le but principal est d'améliorer le taux de reconnaissance en faisant des corrections orthographiques ou morphologiques à l'aide de dictionnaires [17] et de corriger les erreurs de l'étape précédente en s'appuyant sur des informations lexicales, syntaxiques et sémantiques.

Le principal objectif de la phase de post-traitement au niveau lexicale est de déterminer quel est le meilleur mot du lexique pouvant correspondre aux hypothèses de reconnaissance [19].

4. Conclusion

Nous avons présenté dans ce chapitre le domaine de la reconnaissance automatique de l'écriture, les étapes suivies pour la réalisation d'un système de reconnaissance, les différentes distinctions faites entre les systèmes en-ligne et les systèmes hors-lignes, et entre l'écriture imprimée et manuscrite.

Les difficultés rencontrées dans la reconnaissance de l'écriture manuscrite arabe sont de même type que celles affrontées lors de la reconnaissance du latin manuscrit, mais souvent plus complexes à cause de la diversité des formes des caractères arabes, de la courte liaison qui existe entre deux caractères successifs, de l'allongement des ligatures horizontales et de la présence de ligatures verticales.

Le traitement automatique des montants littéraux arabes est un problème délicat à traiter car il regroupe, à lui seul, une grande partie des difficultés généralement rencontrées en reconnaissance de l'écriture manuscrite arabe.

Chapitre3

Le traitement automatique des montants littéraux arabes

Résumé: Dans ce chapitre, nous présenterons les caractéristiques morphologiques de l'écriture arabe. Ensuite nous exposerons les principaux travaux développés en reconnaissance automatique des montants littéraux arabes.

1. Introduction

Le traitement des chèques comporte l'ensemble des tâches qu'un agent de banque peut avoir faire à un nouveau chèque d'un client. Ceci inclut: l'accès aux numéros de compte, la vérification des noms et des signatures sur les chèques, la reconnaissance du montant numérique, la reconnaissance du montant littéral (le seul qui soit légal), la vérification de la date, et la correspondance entre le montant numérique et le montant littéral. La lecture automatique des chèques est donc une application industrielle qui se compose de plusieurs sous applications, dont le champ de la recherche est rapide et conviviale. Ces applications nécessitent de faibles taux de rejets et taux d'erreurs car le coût des erreurs est extrêmement élevé.

Plusieurs produits commerciaux sont proposés pour la lecture automatique des chèques (les principaux produits ou prototypes disponibles sur le marché français: le SRTP (La Poste) avec Dassault AT (Automate de remise de chèques lisant le montant numérique seulement) ; Matra (MSI) et son produit LireChèques ; Itesoft et son produit ChequeScan, et enfin le produit CheckReader de la société A2IA. En 2007, A2iA lance l'Auto-balancing, une fonctionnalité innovante de son logiciel A2iA CheckReader qui augmente le taux de reconnaissance des remises de chèques et qui améliore les performances des solutions de ses partenaires intégrateurs. A2iA CheckReader, intégrant l'Auto-balancing, est disponible en 6 langues (français, anglais, allemand, espagnol, italien et portugais) et 20 versions pays (France, Grande-Bretagne, Irlande, Italie, USA, Canada, Mexique, Brésil, Thaïlande, Hong Kong, Suède, Pays-Bas, Portugal, Nouvelle-Calédonie, Australie, Allemagne, Chili, Belgique, Malaisie, Singapour)[31].

L'Arabe a un plus grand vocabulaire au niveau du montant littéral que l'anglais et le français. Le grand nombre des composantes secondaires (les points diacritiques, le hamza) cause plus de difficultés au niveau de la segmentation. En outre, les règles grammaticales de l'Arabe permettent de grandes variations dans l'écriture des montants littéraux (voir figure 3.1) [32]. Par conséquent, il est nécessaire de construire une base de données arabe afin d'apprendre, de tester et de comparer les systèmes de reconnaissance.

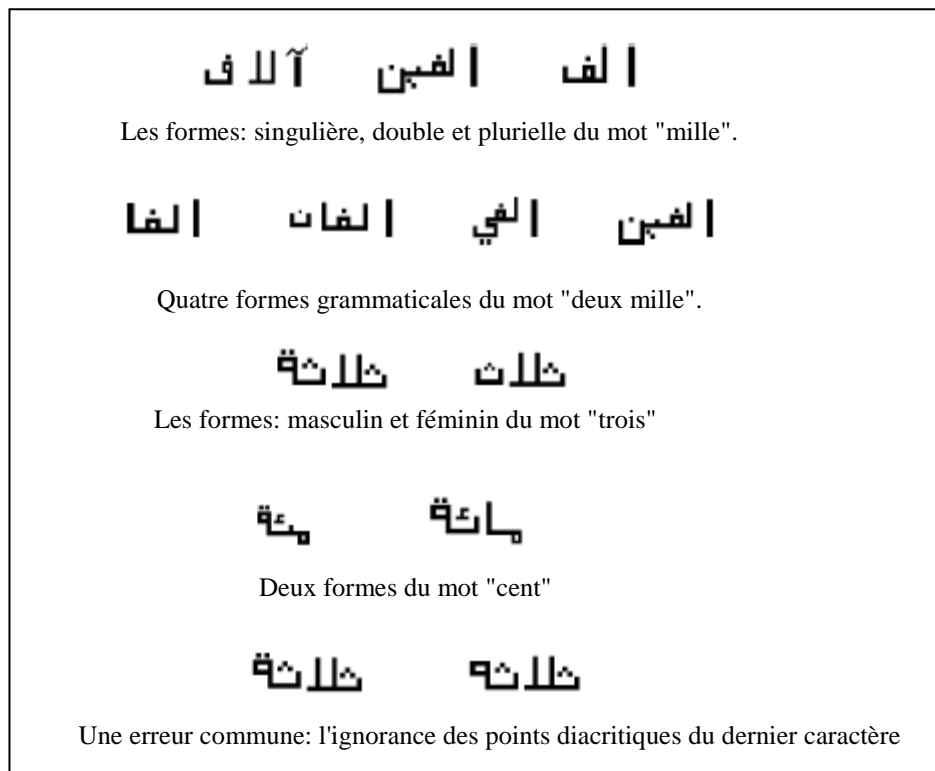


Figure 3.1 : des variations dans l'écriture des montants littéraux [32]

2. Caractéristiques morphologiques de l'écriture arabe

L'Arabe est écrit par plus de cent millions de gens, dans plus de vingt pays différents. L'écriture arabe a été développée à partir d'un type d'Araméen. La langue araméenne comporte moins de consonants que l'arabe, alors de nouvelles lettres ont été créées en ajoutant des points aux lettres déjà existantes. D'autres petites marques appelées diacritiques sont utilisées pour indiquer de courtes voyelles, mais elles ne sont généralement pas utilisées [33].

L'alphabet arabe

L'alphabet arabe comporte 28 lettres (consonnes) et la Hamza « ء » qui est le plus souvent considérée comme signe complémentaire, et le symbole « ~ » qui s'écrit uniquement

sur le support du caractère « ٰ », l'alphabet arabe comprend d'autres caractères additionnels tels que « ة » et « لا »,

Les caractères arabes s'écrivent de façon cursive, de droite vers la gauche, aussi bien dans le cas de l'imprimé que du manuscrit, et certains caractères arabes incluent une boucle qui peut avoir différentes formes (voir figure 3.2).

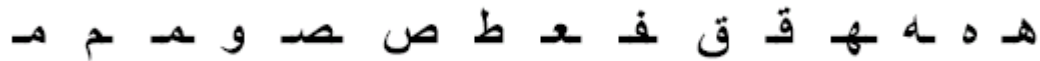


Figure 3.2 : Exemple de différentes formes de la boucle

Les caractéristiques de l'écriture arabe

La forme d'un caractère diffère selon sa position dans le pseudo-mot et même dans certains cas, selon le contexte phonétique. En outre, plus de la moitié des caractères arabes incluent dans leur forme des points diacritiques (1,2 ou 3). Ces points peuvent se situer au-dessus ou au-dessous du caractère, mais jamais en haut et en bas simultanément. Plusieurs caractères peuvent avoir le même corps mais un nombre et /ou une position de points diacritiques différents [34, 17].

Pour des raisons de justification de texte et/ou d'esthétique, les ligatures horizontales peuvent être allongées en insérant entre les caractères d'une même chaîne une ou plusieurs élancements «matta» (ou tatwil), correspondant au symbole «←». L'élancement se situe toujours à gauche du caractère courant. Si le trait d'allongement est associé à un caractère en position de début ou finale, le caractère prend sa forme de milieu et voit sa chasse augmenter du nombre de « matta » insérées (tableau 3.1) [35].

Certains caractères ne peuvent être rattachés à leur gauche et de ce fait ne peuvent se trouver qu'en position isolée ou finale, ce qui donne quand ils existent, des mots composés d'une ou de plusieurs parties qu'il est convenu d'appeler généralement PAW (piece of arabic word) ou encore pseudo-mot [17] ou sous-mot. Un PAW correspond donc à une chaîne d'un ou de plusieurs caractères. L'écriture arabe est ainsi semi-cursive plutôt que totalement cursive.

De plus, certains caractères peuvent être écrits de façon combinée. Ces combinaisons ou ligatures, sont optionnelles contrairement aux ligatures horizontales qui sont obligatoires [13].

Il existe 8 signes de voyellation qui peuvent se placer au dessus de la ligne d'écriture, tels que fathah ($\overset{\sim}{\text{—}}$) dhammah ($\overset{\text{ٓ}}{\text{—}}$), soukoun ($\overset{\circ}{\text{—}}$) et chaddah ($\overset{\text{و}}{\text{—}}$) qui doit être accompagnée de l'une des voyellations fatha, Dammah ou kasrah, en dessous tels que Kasrah ($\overset{\sim}{\text{—}}$). De plus trois «tanwin » peuvent être formés à partir d'un double fatha ($\overset{\sim}{\text{—}}$), d'un double dhammah ($\overset{\text{ٓ}}{\text{—}}$) ou d'un double kasrah ($\overset{\sim}{\text{—}}$).

caractère	initiale	médiane	finale	Isolé
Alif			ا	ا
Ba	ب	ب	ب	ب
Ta	ت	ت	ت	ت
Tha	ث	ث	ث	ث
Jim	ج	ج	ج	ج
Ha	ح	ح	ح	ح
Kha	خ	خ	خ	خ
Dal			د	د
Thal			ذ	ذ
Ra			ر	ر
Zay			ز	ز
Sin	س	س	س	س
Chin	ش	ش	ش	ش
Sad	ص	ص	ص	ص
Dhad	ض	ض	ض	ض
Tad	ط	ط	ط	ط
Dha	ظ	ظ	ظ	ظ
Ayn	ع	ع	ع	ع
Ghayn	غ	غ	غ	غ
Fa	ف	ف	ف	ف
Qaf	ق	ق	ق	ق
Kaf	ك	ك	ك	ك
Lam	ل	ل	ل	ل
Mim	م	م	م	م
Noun	ن	ن	ن	ن
He	ه	ه	ه	ه
Waw			و	و
Ya	ي	ي	ي	ي

(a)

caractère	initiale	médiane	finale	isolé
Ta			ة	ة
Lamalif			لا	لا

(b)

caractère	initiale	médiane	finale	isolé
Alif+~			أ	أ
Alif+ء			آ	آ
			إ	إ
Waw+ء			ؤ	ؤ
Ya+ء	ئ	ئ	ى	ى

(c)

caractère	initiale	médiane	finale	isolé
Lamalif +~			لا	لا
Lamalif +ء			لا	لا
			لا	لا

(d)

Tableau 3.1:

- (a) – l’alphabet arabe dans ses différentes formes.
- (b) – les caractères additionnels
- (c), (d) – Hamza et Madda et les positions qu’elles occupent en association avec Alif, Waw et Ya [36].

3. Quelques systèmes de traitement automatique des montants littéraux arabes

Pour concevoir et améliorer les performances d'un système de reconnaissance de l'écriture manuscrite en générale et en arabe en particulier pour le traitement automatique des chèques, il faut bien choisir le paradigme (approche globale, analytique ou hybride), et les techniques de reconnaissances (Réseaux de neurones, SVM, MMC, Symbolique, Système Expert, Logique floue, Syntaxique, Structurelle, Statistiques, ou Hybride).

Dans ce qui suit, une synthèse de quelques systèmes de traitement automatique des montants littéraux arabes basés sur différents paradigmes et techniques de reconnaissance sera présentée.

Le système de Al-Ohli, 2002 [32]

Il a adopté un outil de marquage conçu pour marquer des dates latines, qui peut être réutilisé pour marquer les sous-mots et les chiffres arabes. Le marquage essaie d'étiqueter chaque objet du montant littéral ou du montant numérique dans une image prétraitée.

L'approche utilisée est l'approche hybride où l'entité à reconnaître est le sous-mot, dans le tableau 3.2, l'auteur donne la distribution des sous-mots dans la base d'image.

Le système utilise pour la reconnaissance la technique stochastique basée sur les chaînes de Markov cachées (MMC), pour fournir des mesures probabilistes fondées sur des observations séquentielles. Ces dernières sont détectées à partir de l'entrée image 2D basé sur une estimation des séquences de segments. Pour faciliter l'extraction des caractéristiques analytiques, la squelettisation est appliquée à l'entrée de l'image de sous-mot.

Avant l'extraction des primitives les plus pertinentes, un prétraitement sur l'image est effectué en deux étapes : segmentation statique et seuillage dynamique consiste à déterminer automatiquement la séparation entre le fond et la forme (voir figure 3.3).

Code	Image	Count	Code	Image	Count	Code	Image	Count	Code	Image	Count	Code	Image	Count
1-00	ا	2822	2-09	لا	2450	3-16	سنا	1	6-06	هالين	0	4-23	هتني	0
1-05	ت	6	2-10	لفا	1049	3-17	سقة	160	4-01	بعيا	118	4-24	هتعة	1
1-06	ن	14	2-11	يا	299	3-19	ستو	89	4-02	بعين	8	4-26	هاللا	0
1-07	د	32	2-12	نوو	94	3-20	سعو	34	4-03	تسما	0	4-27	هاللا	6
1-08	ر	2920	2-13	يا	1830	3-21	عشر	671	4-04	صمعة	90	5-00	هتعة	1
1-09	ف	811	2-14	يين	16	3-22	غير	1519	4-06	شمعو	58	5-02	صميا	116
1-10	ل	1665	3-00	عنا	71	3-23	فقط	1415	4-07	هتعة	0	5-03	شمين	1
1-11	ن	1357	3-02	هتني	4	3-24	لغا	179	4-09	شمين	18	5-04	هتيا	263
1-12	ة	345	3-04	هتعة	222	3-25	لغا	17	4-10	هتيا	1	5-05	هتين	14
1-14	و	2677	3-06	هتو	117	3-27	لغي	3	4-11	هتعة	392	5-06	هتيا	118
1-16	ي	53	3-07	هتس	5	3-28	لين	1	4-13	هتسو	213	5-07	هتين	1
2-00	هتعة	1152	3-08	هتلا	509	3-29	هتعة	14	4-15	هتسعة	121	5-08	هتعة	1
2-02	هتيم	9	3-09	هتيا	440	3-31	هتيا	99	4-17	هتسو	102	5-10	هتشمين	4
2-03	هتشة	269	3-10	هتعا	77	3-33	هتية	132	4-18	هتيا	132	5-11	هتلا	0
2-05	هتوو	128	3-12	هتني	20	3-35	هتين	2	4-19	هتشمين	5	6-00	هتصمعة	1
2-06	هتفا	93	3-13	هتشمين	6	3-36	هتلا	2	4-20	هتشمين	89	6-02	هتصمعة	1
2-07	هتستا	3	3-14	هتشمين	15	4-00	هتشمين	31	4-21	هتعا	48	6-04	هتصمعة	1
2-08	هتقا	111	3-15	هتشم	2									

Tableau 3.2 : la distribution de sous-mots dans la base d'image [32]

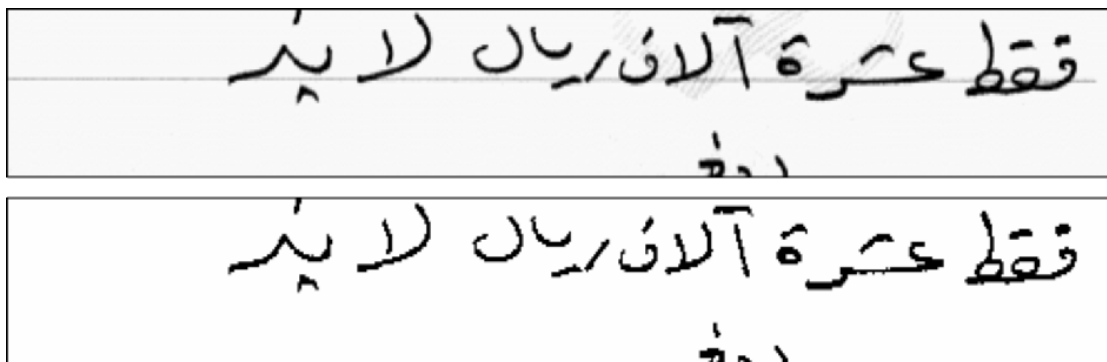


Figure 3.3 : le prétraitement [32]

Le classifieur markovien, basé sur les caractéristiques locales du sous-mot, et les modèles sont entraînés en utilisant l’algorithme de Baum-Welch.

La collaboration entre le CENPARMI (Centre For Pattern Recognition and Machine Intelligence) et LIVIA (Laboratoire d'Imagerie, Vision et d'Intelligence Artificielle, École de Technologie Supérieure, Université du Québec) d’une part et le Banque d’Al Rajhi Banking and

Investment Corporation (Une des plus grandes sociétés bancaires en Arabie Saoudite) de l'autre part, a donné lieu à une collection du monde réel de 7000 images de chèques en niveau de gris avec une résolution de 300dpi (figure 3.4).



Figure 3.4 : image d'un chèque de la base d'images CENPARMI [32]

ثمانون ألف ريال سعودي

Figure 3.5 : le montant littéral [32]

L'auteur utilise dans l'étape de l'apprentissage 20 879 sous-mots et 1779 montants littéraux, et 8619 sous-mots et 720 montants littéraux sont réservés pour l'étape de test. Le taux de reconnaissance a atteint 73%.

Le système de Al-Ma'adeed, 2004 [37]

Le système de [37] est un système de reconnaissance de mots manuscrits, basé sur une approche hybride.

Le prétraitement comporte essentiellement la tâche de normalisation pour chacun des facteurs suivants: la largeur du trait, la pente et la hauteur des lettres.

La classification se fait par combinaison séquentielle de deux classifieurs. Le premier basé sur des règles et utilise trois caractéristiques perceptuelles globales (nombre de points hauts, nombre de points bas et nombre de segments), l'utilisation de ces caractéristiques avec une codification de l'existence un ou zéro implique (génère) 8 groupes, ce classifieur tente

d'affecter le mot à reconnaître au groupe approprié parmi les 8 groupes prédéfinis. Et le second est un classifieur local de type markovien sert à la classification des mots au sein du groupe concerné (voir figure 3.6). Les états du MMC sont identifiés par les lettres de l'alphabet. Une fois le modèle est établi, l'algorithme de Viterbi est utilisé pour faire la reconnaissance des lettres composant le mot.

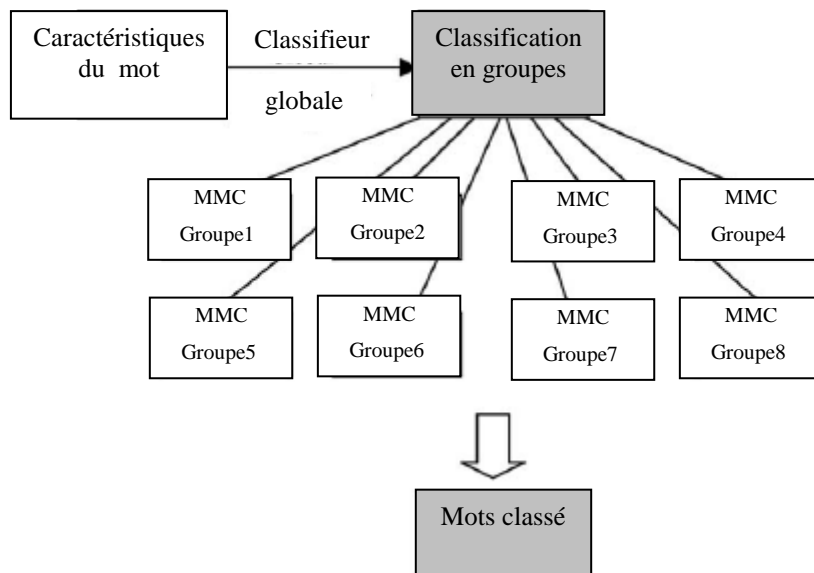


Figure 3.6: processus de reconnaissance en utilisant une combinaison de deux classifieurs [37]

Le système décrit au-dessus a été appliqué à une base d'images de mots manuscrits produits par 100 scripteurs et contient 4700 mots (voir des exemples dans la figure 3.7). Environ 10% des données ont été exclues de test et d'apprentissage à cause d'erreurs dans la détection de la ligne de base et de pré-traitement.

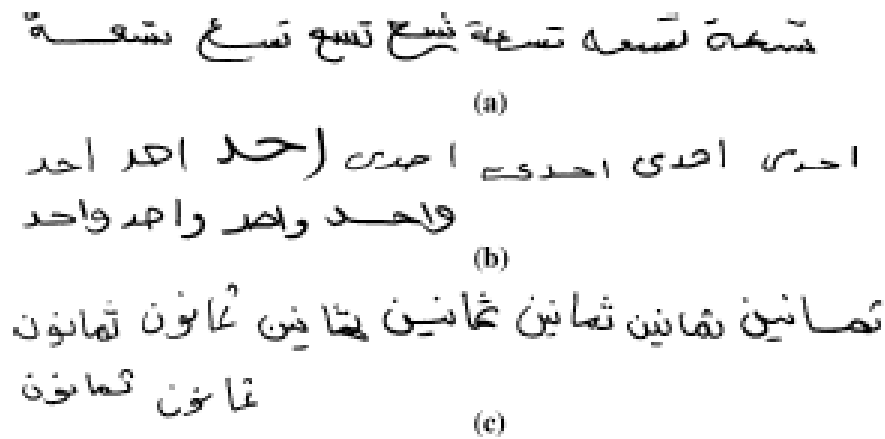


Figure 3.7: Des exemples des mots arabes appartenants aux montants littéraux; (a)- neuf (b)- un ou part de onze (c)- quatre-vingts [37].

Le taux de reconnaissance pour le premier classifieur a atteint 99% (tableau 3.3), mais pour le système complet, le taux de reconnaissance a avoisiné 60% (tableau 3.4).

Caractéristiques utilisées	Taux de reconnaissance en groupes		
	(Choisir le meilleur taux)(%)	(Choisir les deux meilleurs taux) (%)	(Choisir tous les taux) (%)
1, 2, 3	94.86	98.46	99.10

Tableau 3.3: taux de reconnaissance du classifieur globale [37]

Gr1	Gr2	Gr3	Gr4	Gr5	Gr6	Gr7	Gr8	Taux de reconnaissance	T. de rec. avec une simple grammaire
97%	55%	55%	78%	100%	56%	59%	57%	60%	69%

Tableau 3.4: taux de reconnaissance du classifieur local et du système total [37]

Le système de N. Farah 2005 [38]

Le système de [38] est un système de reconnaissance de montants littéraux arabes, basé sur une approche globale, il utilise une combinaison de trois classifieurs (réseau neuronal de type perceptron multi-couches, k plus proches voisins et k plus proches voisins flou).

Avant l'extraction de caractéristiques, les données du système sont subies à un prétraitement comporte la segmentation (l'espace inter-mot est 1.5 fois plus grand que l'espace inter-sous-mots), la binarisation, la squelettisation et le redressement de la ligne de base.

Les trois classifieurs exploitent les mêmes caractéristiques globales structurelles (les ascendants, les descendants, les boucles, les simples points hauts, les simples points bas, les doubles points hauts, les doubles points bas, les triples points hauts et les sous-mots).

La figure 3.8 illustre le processus de reconnaissance du système.

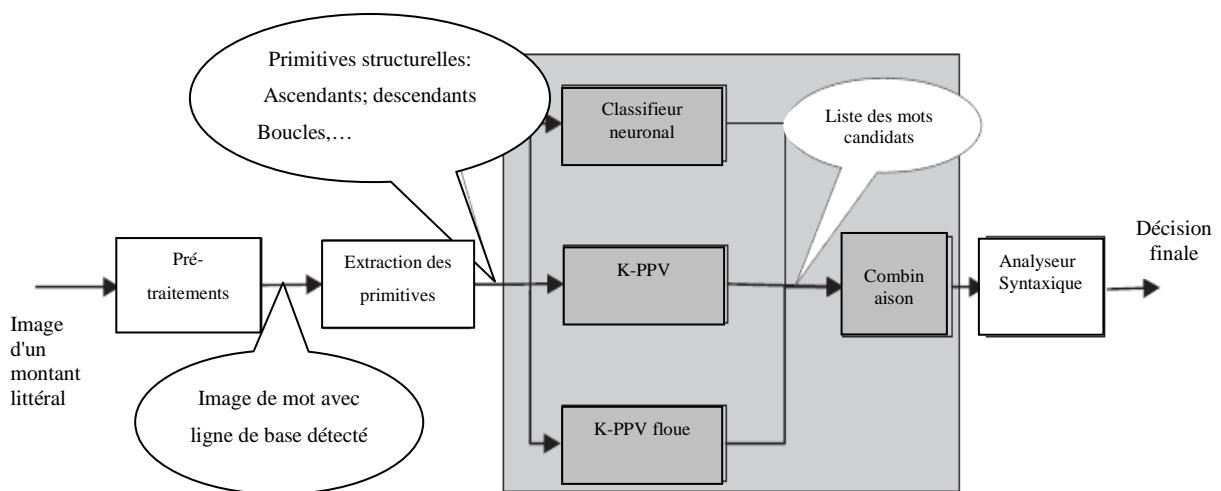


Figure 3.8: le processus du système de reconnaissance [38]

Le premier classifieur neuronal repose sur les caractéristiques globales structurelles d'un mot présenté sous forme d'une image à l'entrée de ce système. Le réseau de neurones utilisé est le perceptron multicouche (PMC) avec un apprentissage supervisé (optimisation des valeurs de poids).

- la couche d'entrée: Le réseau a comme des entrées les primitives structurelles (21 neurones pour 21 primitives, par exemples: 3 neurones pour les ascendants, 2 neurones pour double points bas.
- la couche de sortie: se compose de 48 neurones représentant les mots de vocabulaire des montants littéraux arabes.
- une couche cachées: le nombre de neurones est calculé heuristiquement (la racine carré du nombre des entrées * nombre des sorties).

Le second classifieur k plus proches voisins, basé sur les mêmes caractéristiques. On cherche les k individus les plus proches entre le vecteur de caractéristiques du mot testé et les vecteurs de caractéristiques de l'ensemble d'apprentissage (en utilisant les distances euclidiennes).

Le troisième classifieur k plus proches voisins flous: basé aussi sur les mêmes caractéristiques. On calcul l'adhésion d'un mot testé aux différentes classes et on l'affecte à la classe qui a la valeur le plus grande.

La combinaison des classifieurs

Il y a plusieurs stratégies de combinaison des classifieurs ; on peut sommer ou multiplier les valeurs de confiances accordées aux réponses des classifieurs ; on peut aussi prendre les valeurs maximum / minimum.

Chaque classifieur donne en sortie une liste, cette liste comporte les trois meilleurs réponses avec ses valeurs de confiance ; ces valeurs de confiance sont normalisées pour correspondre aux $P(w_i|x)$ (probabilité à posteriori).

La post-classification syntaxique : est le responsable de vérifier la solution correcte au lieu des solutions retournées par les classifieurs et de la sélectionner en utilisant un haut niveau d'information qui n'est pas disponible lors de l'étape de la classification.

La base d'image utilisée contient 4800 mots, représentant les 48 mots du lexique écrit par 100 auteurs. 1200 mots ont été utilisés durant l'étape d'apprentissage, et 3600 mots ont été exploités dans l'étape de test.

Les résultats de reconnaissance avant combinaison sont présentés dans le tableau 3.5.

Le classifieur	RNA	K-PPV	K-PPVf
Le taux	91%	89.08%	92.16%

Tableau 3.5: taux de reconnaissance de chaque classifieur [38]

La stratégie de combinaison	Le taux de reconnaissance (%)
Minimum	92.10
Maximum	93.90
Moyen	93.40
Produit	92.40
Somme	94.00

Tableau 3.6: *taux de reconnaissance de la combinaison* [38]

La post-classification augmente le taux de reconnaissance à 96%, cela est dû à l'intégration de l'information contextuelle, qui n'est pas disponible dans la classification (voir le tableau 3.6).

Le système de Souici 2005 [39]

L'approche utilisée dans [39] est une approche hybride neuro-symbolique. L'idée générale de cette approche est d'intégrer des connaissances symboliques au sein d'un réseau neuronal afin d'éviter d'une part les principaux problèmes de l'approche connexionniste, et d'autre part d'essayer de rapprocher, le plus possible, des mécanismes de perception humain (système de reconnaissance visuelle des mots).

Le réseau de neurones utilisé est de type Perceptron multicouches PMC, et la représentation symbolique de la connaissance est basée sur l'hypothèse que l'esprit humain, est une machine de série de traitements de l'information en utilisant des règles de production.

La phase de prétraitement comporte la binarisation, l'adoucissement (smoothing) d'image, la normalisation, la détection de ligne de base et l'extraction de contour.

Les caractéristiques extraites de données sont : Le codage de contour (l'utilisation du code freeman 8-voisins), le nombre des boucles, le nombre des points diacritiques, le nombre des ascendants, et le nombre des descendants, pour le classifieur neuro-symbolique (voir figure 3.9).

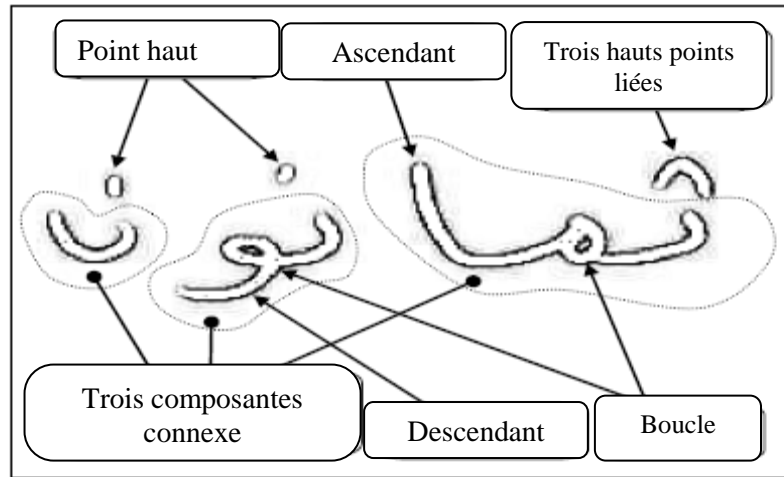


Figure 3.9: Exemple d'extraction de primitives globale [39]

Et la figure 3.10, illustre le processus d'extraction des primitives.

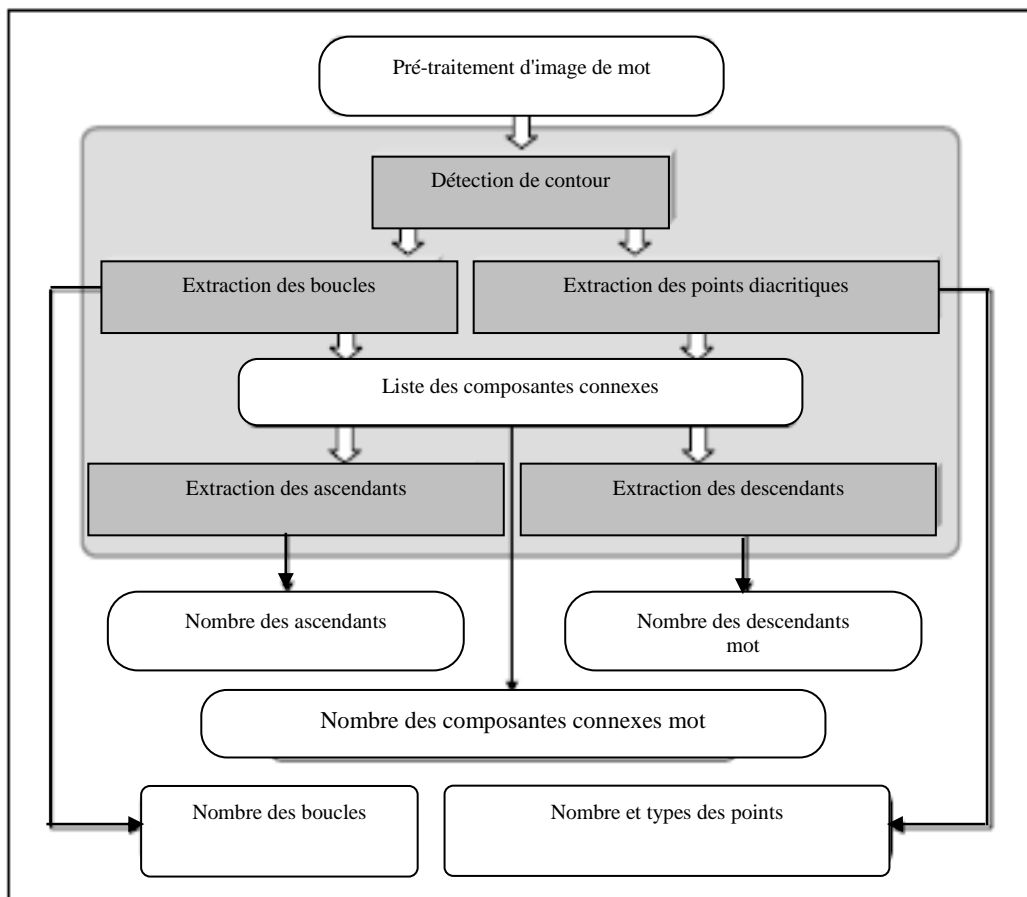


Figure 3.10 : Processus d'extraction des primitives [39]

Le classifieur est inspiré de l'approche KBANN (Knowledge Based Artificial Neural Network) développé par Towell en 1991 et testé dans le domaine de la biologie moléculaire...etc

Les connaissances symboliques initiales (connaissances théoriques) modélisées par les règles sont intégrées dans un réseau de neurones en définissant sa topologie et ses poids initiaux des liens. Le réseau est ensuite affiné au moyen d'un algorithme d'apprentissage (rétro-propagation) sur une base d'apprentissage (voir figure 3.11).

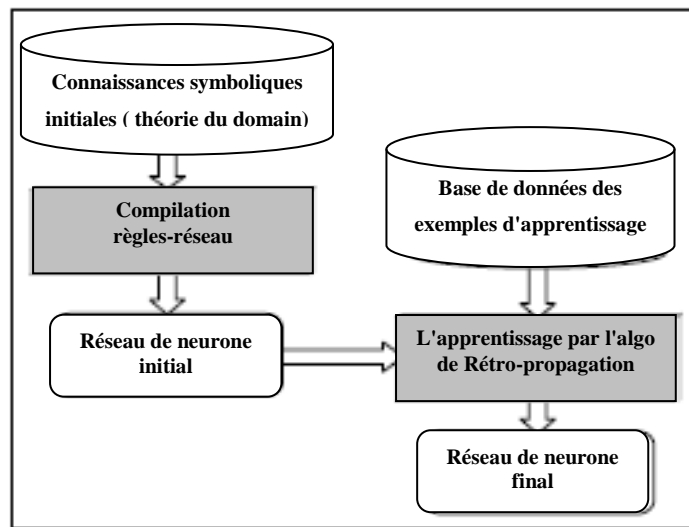


Figure 3.11 :L'intégration neuro-symbolique [39]

Il est donc nécessaire de suivre les étapes suivantes pour concevoir le classifieur selon l'approche neuro-symbolique:

- construire une base de connaissances décrivant les mots arabes avec leurs caractéristiques (caractéristiques structurelles décrites les 48 mots du vocabulaire des montants littéraux arabes). L'ensemble de règles (de la forme: si – caractéristique → alors –conclusion) doit être hiérarchiquement structuré.
- traduire les règles en un réseau de neurones avec un algorithme de compilation règles-réseau.
- raffiner le réseau de neurones à l'aide d'un processus d'apprentissage empirique (algorithme de rétro-propagation).

La base d'image utilisée contient 576 mots (les 48 mots du vocabulaire ont été écrits trois fois par quatre scripteurs) pour l'apprentissage, et 432 mots (les 48 mots du vocabulaire

ont été écrits trois fois par trois scripteurs) pour le test. D'autres bases de données indépendantes pour l'apprentissage et le test ; la première contient 480 mots (48 mots écrits par 10 scripteurs) et la seconde contient 1200 mots (48 mots écrits par 25 scripteurs).

Le taux de reconnaissance est donné dans le tableau 3.7 suivant :

Type de classifieur	Taux de reconnaissance (%)
Symbolique (à base de règles)	83.55
Réseau de neurones	85.50
Hybride neuro-symbolique	93.00

Tableau 3.7 : taux de reconnaissance de chaque approche et d'hybridation [39]

4. Conclusion

La problématique associée à la reconnaissance des montants littéraux arabes de chèques bancaires et postaux est compliquée et plus riche à la fois. Les travaux précédemment présentés sont arrivés à dépasser plusieurs obstacles et ont un développement rapide et des résultats encourageants. Mais il reste beaucoup de travail à faire sur ce sujet pour pallier des problèmes spécifiques aux deux approches: globale (entité à reconnaître est le mot) et pseudo-analytique (entité à reconnaître est le sous-mot) tels que :

- les boucles ne sont pas clairement distinguables.
- chevauchement vertical des caractères change leurs morphologies.
- chevauchement horizontale et l'absence des ligatures horizontales qui impliquent une séparation incorrecte des sous-mots (Ce sont les problèmes les plus fréquemment rencontrés avec les composantes connexes).
- l'ignorance des points diacritiques.

Chapitre4

Les modèles de Markov cachés (MMC)

***Résumé:** Dans ce chapitre, nous présenterons les modèles de Markov cachés MMCs. Tout d'abord nous commencerons par une introduction aux MMCs, et une présentation de quelques définitions des concepts qui sont en relation avec les MMCs et nous donnerons aussi les éléments qui constituent un MMCs. Ensuite nous exposerons les trois problèmes fondamentaux des MMCs, Et pour finir nous montrerons l'utilisation de ces modèles pour la reconnaissance automatique de l'écriture.*

1. Introduction

La modélisation stochastique permet l'utilisation des modèles probabilistes pour traiter les problèmes à information incertaine ou incomplète. Ainsi, les modèles de Markov connaissent un regain d'intérêt tant dans leurs aspects théoriques qu'appliqués.

Les modèles de Markov cachés MMCs (HMM, pour Hidden Markov Model) sont des méthodes de modélisation stochastiques introduites fin des années 60, début des années 70 pour de nombreuses applications, principalement dans le domaine de la parole, pour l'analyse de l'ADN, la prédiction de l'évolution des épidémies,.....

Les modèles de Markov connaissent actuellement un essor important en reconnaissance des formes grâce à leur capacité d'intégration du contexte et d'absorption du bruit [40, 16].

L'utilisation des MMCs en reconnaissance automatique de l'écriture a permis d'obtenir des résultats intéressants pour certaines applications. Les différents travaux réalisés reposent pour une grande part sur l'expérience accumulée dans le domaine de la reconnaissance de la parole où les MMCs sont fréquemment utilisés. Comparés à d'autres approches de reconnaissance (structurale, géométrique, etc.), les MMCs se distinguent par leur capacité de modéliser efficacement différentes sources de connaissance. En effet, d'une part ils offrent une intégration cohérente de différents niveaux de modélisation (morphologique, lexical et syntaxique) et d'autre part, il existe des algorithmes puissants permettant de déterminer la

valeur optimale des paramètres fournissant la meilleure adéquation entre le modèle et la base de données (connue) qualifiée d'apprentissage [41, 34].

2. Définition des MMCs

Un modèle de Markov caché est un double processus stochastique (X_t, Y_t) $1 \leq t \leq T$. La chaîne interne X_t non observable, et la chaîne externe Y_t observable, s'allient pour générer le processus stochastique. La chaîne interne est supposée, pour chaque instant, être dans un état où la fonction correspondante génère une composante de l'observation. La chaîne interne change d'état en suivant une loi de transition. L'observateur ne peut voir que les sorties des fonctions aléatoires associées aux états et ne peut pas observer les états de la chaîne sous-jacente, d'où le terme de Modèles de Markov Cachés (ou Hidden Markov Model) [42].

Le processus (X_t) $0 \leq t \leq T$ est une chaîne de Markov d'ordre 1, il doit vérifier :

$$P(X_{t+1} = q_j / X_t = q_i, \dots, X_0 = q_0) = P(X_{t+1} = q_j / X_t = q_i) = a_{ij} \text{ pour tout } t \geq 0. \quad (4.1)$$

Le processus (Y_t) $0 \leq t \leq T$, processus observable, vérifie :

$$P(Y_t = y_t / X_t = q_i, \dots, X_1 = q_1, Y_{t-1} = y_{t-1}, \dots, Y_1 = y_1) = P(Y_t = y_t / X_t = q_i) = b_i(y_t) \quad (4.2)$$

Les observations sont supposées indépendantes les unes des autres conditionnellement à la suite d'états. Chaque réalisation de Y_t ne dépend que de l'état courant caché. Les observations y_t peuvent être de nature [42]:

- **discrète** $\Rightarrow b_i$ est une distribution de probabilité discrète : une loi discrète est généralement représentée par les fréquences d'apparitions des observations discrètes.
- **continue** $\Rightarrow b_i$ est une fonction de densité de probabilité définie sur \mathbb{R}^d : les densités traditionnelles utilisées sont des densités gaussiennes, entièrement définies par le vecteur moyen et la matrice de covariance, ou des densités de type multi-gaussiennes (sommées pondérées de densités gaussiennes).

Et pour mieux comprendre, voici un exemple théorique extrait de [43]. Considérant deux personnes séparées par un mur. La première personne possède trois dés biaisés. Tour par tour, elle choisit un dé, le lance et énonce à voix haute le numéro de la face résultante à la deuxième personne. La deuxième personne de l'autre côté du mur, ne connaît que la séquence de numéros de dés, pas la séquence des dés qui a été choisie par la première personne. La séquence des dés est appelée séquence des états cachés, et la séquence des numéros de face est appelée séquence d'observations.

3. Les éléments d'un MMC

Selon [40], un MMC est caractérisé par :

- N , le nombre d'états dans le modèle. On note $S = \{s_1, s_2, \dots, s_N\}$, l'ensemble des N états du modèle, q_t l'état au temps t (q_t appartient à S).
- M , le nombre de symboles d'observation par état, c-à-dire la taille de l'alphabet discret, les symboles d'observation correspondent aux sorties physiques du système à modéliser. On note $V = \{v_1, v_2, \dots, v_M\}$, l'ensemble discret des M symboles, et o_t un symbole au temps t (o_t appartient à V).
- La distribution de probabilité de transition d'une état $A = \{a_{ij}\}$. Tel que :

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], \quad 1 \leq i, j \leq N. \quad (4.3)$$

Pour le cas spécial où à partir de tout état, on peut atteindre directement n'importe quel autre état, on a : $a_{ij} > 0$, pour tout i, j . Pour autre types d' MMC, on a $a_{ij} = 0$ pour un seul pair (i, j) ou plus.

- La distribution de probabilité de symboles d'observation dans un état j , $B = \{b_j(k)\}$, où :

$$b_j(k) = P[v_k \text{ en } t | q_t = S_j], \quad 1 \leq j \leq N, \quad 1 \leq k \leq M. \quad (4.4)$$

- La distribution d'état initial $\pi = \{\pi_i\}$ tel que :

$$\pi_i = P[q_1 = s_i], \quad 1 \leq i \leq N. \quad (4.5)$$

Donc, la spécification complète d'un MMC nécessite la spécification des deux paramètres du modèle N et M , la spécification des symboles d'observations, et la spécification des trois mesures de probabilités A , B et π . L'ensemble complet des paramètres d'un MMC peut être noté : $\lambda = (A, B, \pi)$.

Dans un MMC, les contraintes (markoviennes) suivantes doivent être respectées :

$$\sum_{i=1}^N \pi_i = 1, \quad (4.6)$$

$$\sum_{j=1}^N a_{ij} = 1, \quad 1 \leq i \leq N, \quad (4.7)$$

$$\sum_{k=1}^M b_j(k) = 1, \quad 1 \leq j \leq N \quad (4.8)$$

4. Extensions des MMCs

densités de probabilités

Selon le type de densité de probabilité d'observations, discrète ou continue, il est possible de construire deux types de modèles de MMC: soit un MMC discret soit un MMC continu [44].

- **MMC discret «Discret Hidden Markov Models (DHMM) »** : les observations en général sont continues puisqu'elles proviennent de phénomènes physiques continus. Dans le cas d'un MMC discret, les observations continues sont quantifiées à l'aide d'un dictionnaire (l'alphabet d'observations dénombrables).

- **MMC continu « Continuous Hidden Markov Models (CHMM) »** : bien qu'il soit possible de quantifier les observations continues, il peut y avoir une sérieuse dégradation d'information associée à cette quantification. Il sera, alors, avantageux de choisir une fonction de densité de probabilités d'observations continues, conditionnée par les états du processus.

Voici un tableau comparatif de ces deux modèles :

	A. MMC continu	B. MMC discret
Nombre de paramètres à estimer	Un nombre élevé	Moins que A
Précision de la classification	précis	Moins précis que A
Hypothèses sur la nature des observations	importantes	Moins importantes que A
Implémentation	Difficile et lente	Plus facile et plus rapide que A
Nombre de corpus d'apprentissage	moyen	Plus élevé que A

Tableau 4.1 : comparaison entre DHMM et CHMM [44]

Duré de séjour dans un état

L'un des inconvénients des MMC de base est le manque d'informations concernant la variabilité de la durée de séjour dans un état en favorisant les courtes durées. Pourtant, c'est d'une importance majeure dans certains processus physiques, à titre d'exemple, la variabilité de la durée des sons dans la parole.

Principalement deux méthodes ont été développées :

- la méthode de **Ferguson** basée sur un MMC de Durée variable Discrète « Variable Duration Hidden Markov Model VDHMM »

- et celle de **Levinson** basée sur un MMC de durée variable continue « Continuous Duration Hidden Model CDHMM » [44].

Ordre d'une chaîne

Une limitation des Modèles de Markov Cachés de base est dans sa définition. Ils modélisent un processus markovien. C'est-à-dire que son évolution ne dépend pas de son passé, mais uniquement de son état présent. Le processus est d'ordre un, ce qui n'est pas le cas de nombreuses applications. Dans un processus, si l'état futur dépend des k états précédents, la chaîne est d'ordre k .

Une application à la reconnaissance de l'écriture a été faite par Kendu et Bahl; ils ont montré l'avantage des MMC du second ordre sur ceux du premier ordre. Ils ont également signalé la difficulté d'implémentation des MMC du second ordre due à l'indisponibilité de leurs formules de réestimation dans la littérature.

Ceci a été corrigé par He et Krouille. Ils ont développé et traité une technique d'extension des algorithmes de Viterbi et de Baum-Welch pour un MMC du second ordre et d'ordre quelconque [44].

5. Les types des MMCs

Selon sa topologie, un MMC peut être l'un des deux types, ergodique ou gauche-droite.

MMC ergodique

Dans ce type, tout état est directement atteignable depuis tout autre état. Il est plus général et intéressant lorsque le modèle représente un processus dont on veut suivre les évolutions des états [40].

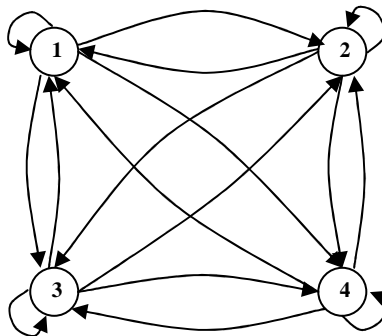


Figure 4.1 : un MMC ergodique [16].

Formellement, on peut écrire :

$$\pi_i \neq 0, a_{ij} \neq 0, \forall i, j \quad (4.9)$$

MMC gauche-droite

Dans ce type, si le temps augmente, alors les indices des états augmentent également. Il est utilisé pour suivre des observations dont l'évolution se fait dans un ordre donné tel que la reconnaissance de la parole [40].

La figure 4.2 suivante illustre deux exemples du MMC gauche-droite (parallèle et séquentielle).

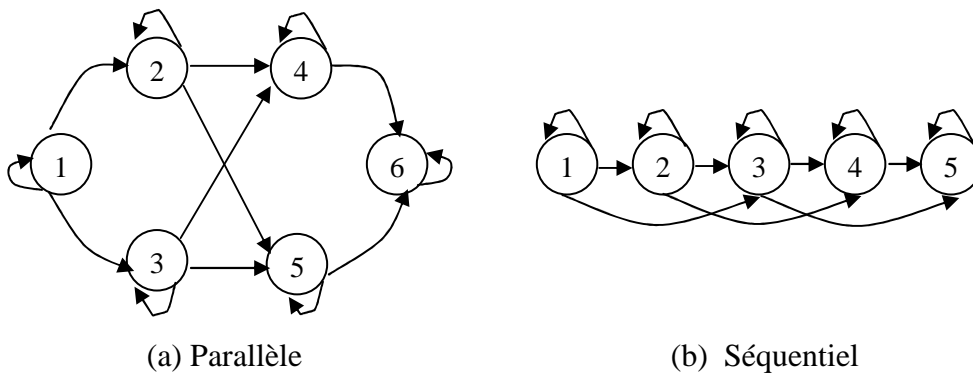


Figure 4.2 : un MMC gauche-droite (a) Parallèle (b) séquentiel [16]

Formellement, on peut écrire :

$$a_{ij} = 0 \text{ si } j < i ; \pi_i = \begin{cases} 0, i \neq 1 \\ 1, i = 1 \end{cases} \quad (4.10)$$

6. Les trois problèmes fondamentaux pour les MMCs

L'utilisation des MMCs nécessite la résolution des trois problèmes fondamentaux suivants :

- problème d'évaluation.
- problème de reconnaissance.
- problème d'apprentissage.

Dans les sections suivantes on va détailler ces problèmes et présenter les algorithmes les plus utilisés pour les résoudre.

Problème d'évaluation

Etant donné une suite d'observations $O = \{o_1, o_2, \dots, o_m\}$ et un modèle λ , comment peut-on calculer efficacement la probabilité (vraisemblance) que la suite d'observations O soit produite par λ , c'est-à-dire $P(O|\lambda)$ [40].

Il existe deux procédures récurrentes de calcul de la vraisemblance [45]:

- l'algorithme Forward-Backward, qui fournit une solution exacte à ce problème faisant intervenir tous les chemins dans le modèle MMC.
- l'algorithme Viterbi, fournissant une solution approximative faisant intervenir uniquement le meilleur chemin dans le modèle MMC.

6.1.1 L'algorithme forward-backward

Dans cette approche, on considère que l'observation peut se faire en deux étapes [40]:

- L'émission de la suite d'observations $\{o_1, o_2, \dots, o_t\}$ et la réalisation de l'état s_t au temps t :

Forward

- L'émission de la suite d'observations $\{o_{t+1}, o_{t+2}, \dots, o_T\}$ en partant de l'état s_t au temps t :

Backward

La 1^{ère} étape suit l'algorithme 4.1 [40]:

- Initialisation :

$$\alpha_1(i) = \pi_i * b_i(o_1), \quad 1 \leq i \leq N,$$

- itération :

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) * a_{ij} \right] * b_j(o_{t+1}), \text{ Pour } 1 \leq t \leq T-1, \quad 1 \leq j \leq N,$$

- terminaison :

$$P(O|\lambda) = \sum_{i=1}^N \alpha_T(i)$$

Algorithme 4.1 : l'algorithme Forward

La figure 4.3 indique comment l'état s_j peut être atteint au temps $t+1$ à partir de N états possibles s_i , $1 \leq i \leq N$, au temps t .

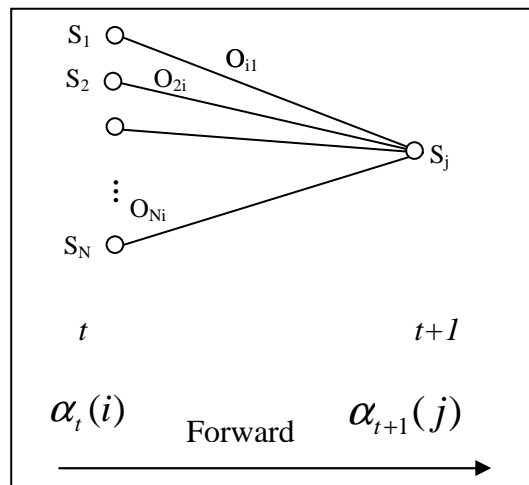


Figure 4.3: Illustration of the séquence of opérations required for the computation of the Forward variable $\alpha_{t+1}(j)$ [40].

La 2^{ième} étape (backward) se déroule comme suit [40]:

On considère une variable backward définie comme :

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = s_i, \lambda), \quad (4.11)$$

C'est-à-dire la probabilité de la séquence d'observation partielle, à partir de $t+1$ jusqu'à la fin, sachant l'état s_i au temps t et le modèle λ (voir figure 4.4).

- initialisation :

$$\beta_T(i) = 1, \quad 1 \leq i \leq N,$$

- itération :

$$\beta_t(i) = \left[\sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \right], \quad \text{pour } 1 \leq t \leq T-1, 1 \leq j \leq N,$$

Algorithme 4.2 : l'algorithme Backward

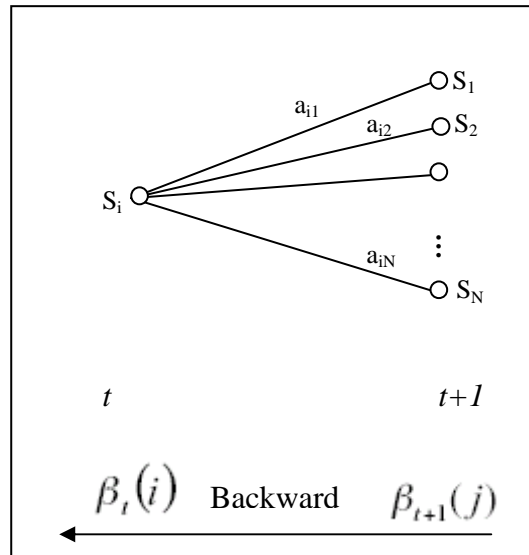


Figure 4.4: Illustration of the séquence of opérations required for the computation of the backward variable $\beta_t(i)$ [40].

6.1.2 L'algorithme de viterbi

Au lieu de prendre en compte tous les chemins autorisés, seul le plus probable est gardé. Ainsi, il suffit de remplacer dans les équations précédentes l'opérateur \sum par \max . Ce critère est largement utilisé du fait du faible coût qui lui est associé (en effet, il est évident que l'opérateur \max est moins coûteux en temps de calcul que l'opérateur \sum sur tous les états). L'algorithme de Viterbi est donc une simplification de la récurrence avant [45].

Problème de reconnaissance

Etant donné un MMC λ et une séquence observée O , comment connaître la séquence des états cachés du λ qui a le plus probablement généré O [40]?

Pour résoudre ce problème la procédure de Viterbi est utilisée. Elle est basée sur les techniques de programmation dynamique. C'est un algorithme récursif qui permet de trouver à partir d'une suite d'observations, une solution optimale au problème d'estimation de la suite d'états.

La procédure complète de l'algorithme de Viterbi pour trouver la meilleure séquence d'états est comme suit:

- initialisation :

$$\delta_1(i) = \pi_i * b_i(o_1) \quad 1 \leq i \leq N,$$

$$\psi_1(i) = 0$$

- calcul récursif :

$$\delta_t(j) = \max_{1 \leq i \leq N} [\delta_{t-1}(i) * a_{ij}] b_i(O_t) \quad 2 \leq t \leq T, 1 \leq j \leq N,$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) * a_{ij}] \quad 2 \leq t \leq T, 1 \leq j \leq N,$$

- terminaison :

$$p^* = \max_{1 \leq i \leq N} [\delta_T(i)]$$

$$q_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$$

- le backtracking :

$$q_t^* = \Psi_{t+1}(q_{t+1}^*), \quad t = T-1, T-2, \dots, 1.$$

Algorithme 4.3 : l'algorithme de Viterbi

Problème d'apprentissage

Etant donné un MMC $\lambda = \{\Pi, A, B\}$ et une séquence observée O , Comment peut-on ajuster les paramètres du modèle λ pour maximiser la vraisemblance $P(O|\lambda)$ [40]?

Les approches les plus utilisées sont basées sur des adaptations de l'algorithme EM (Expectation-Maximisation) appelées [45]:

- algorithme de Baum-Welch : $P(O|\lambda)$ est estimée en tenant compte de tous les chemins possibles (implémentation de l'algorithme Expectation-maximisation (EM)).

- algorithme de Viterbi : $P(O|\lambda)$ est estimée en tenant compte du meilleur chemin uniquement (approximation de l'algorithme EM).

6.3.1 Algorithme de Baum-Welch

L'algorithme Baum-Welch est basé sur le théorème de Baum qui garantit l'atteinte d'un maximum local de la fonction de vraisemblance par ré-estimation des paramètres A, B, π [16]. Cette méthode de Maximum de Vraisemblance est la plus utilisée dans les applications.

Cet algorithme peut être représenté sous la forme itérative suivante [45]:

1. Charger le modèle initial.
 $\lambda_0 = \{ \pi, A, B \}$, N et M
 Avec N : le nombre d'états.
 M : le nombre de symboles distincts.
2. Charger les observations de toutes les occurrences d'un mot.
3. Calculer les nouvelles matrices π , A et B. En utilisant les fonctions Forward $\alpha_t(i)$ et Backward $\beta_t(i)$.
4. Calculer la probabilité totale à partir de la matrice alpha.
5. Retour à l'étape 4, si cette dernière probabilité est supérieure à un seuil fixé déjà et on n'a pas dépassé le nombre d'itération maximal, sinon retour à l'étape 1 pour traiter un autre mot.

Algorithme 4.4 : l'algorithme Baum-Welch

Le test d'arrêt est généralement un nombre d'itérations qui est fixé empiriquement.

Le choix d'un modèle initial influe sur les résultats : toutes les valeurs nulles de A et de B au départ, restent à zéro à la fin de l'apprentissage. Il est à noter que l'algorithme converge vers des valeurs de paramètres qui forment un point critique de $P(O|\lambda)$. Donc, nous obtenons un maximum local ou un point d'inflexion. D'où la nécessité de bien choisir le modèle initial. Pour avoir une estimation convenable du modèle, les ré-estimations se font sur un ensemble de plusieurs suites d'observations appelées corpus d'apprentissage. Donc la taille du corpus d'apprentissage influe, elle aussi, sur les résultats. Il est souhaitable que celle-ci soit importante [44].

6.3.2 Algorithme de Viterbi

Les paramètres sont optimisés de façon à maximiser la vraisemblance du meilleur chemin. Comme pour l'algorithme EM classique, on part d'un ensemble de paramètres initiaux λ_0 et les paramètres optimaux λ sont obtenus de manière itérative. Le processus d'entraînement est composé d'une étape d'estimation **E** qui sert à trouver la segmentation qui maximise la vraisemblance à partir des paramètres, et d'une étape de maximisation (**M**), qui effectue une mise à jour des paramètres étant donné une segmentation. L'ensemble des paramètres initiaux λ_0 peut être estimé à partir de modèles déjà entraînés par un corpus ou par

exemple par l'intermédiaire d'un corpus déjà segmenté. Il est ensuite possible à partir de segmentation optimale trouvée de calculer les paramètres des fonctions de vraisemblance en considérant tous les vecteurs associés à chacune des classes. Ce processus de réaligement des données acoustiques à l'aide d'un modèle et de réentraînement d'un nouveau modèle est effectué jusqu'à ce qu'une certaine convergence soit atteinte (la segmentation ne varie plus ou l'accroissement relatif de la vraisemblance pour l'ensemble des données d'entraînement est inférieur à un seuil fixé) [45].

7. Les MMCs et la reconnaissance automatique de l'écriture

L'écriture est pour nous naturelle, cependant elle est très complexe pour un système de reconnaissance, voir Figure 4.5. Des recherches de A. Kundu et al, ont mené à la modélisation des 26 lettres de l'alphabet par des suites de segments. Le mot « et » est représenté comme ceci [44]:



Figure 4.5 : *Ecriture manuscrite et reconnaissance automatique* [44].

Concrètement (voir figure 4.6), un mot correspond à une image, avant d'être donné en entrée du MMC, il doit être vectorisé, en voici les étapes :

- Segmentation : découpé en petit segment
- Normalisation : ajout des lignes de référence afin de trouver l'orientation du texte
- Extraction des caractéristiques : comparaison aux différents segments connus. Ceci donne la chaîne d'observations du MMC.
- Elle est donnée en entrée aux différents MMC qui représentent les lettres.
- Celui qui aura la plus grande probabilité sera considéré comme le mot reconnu.

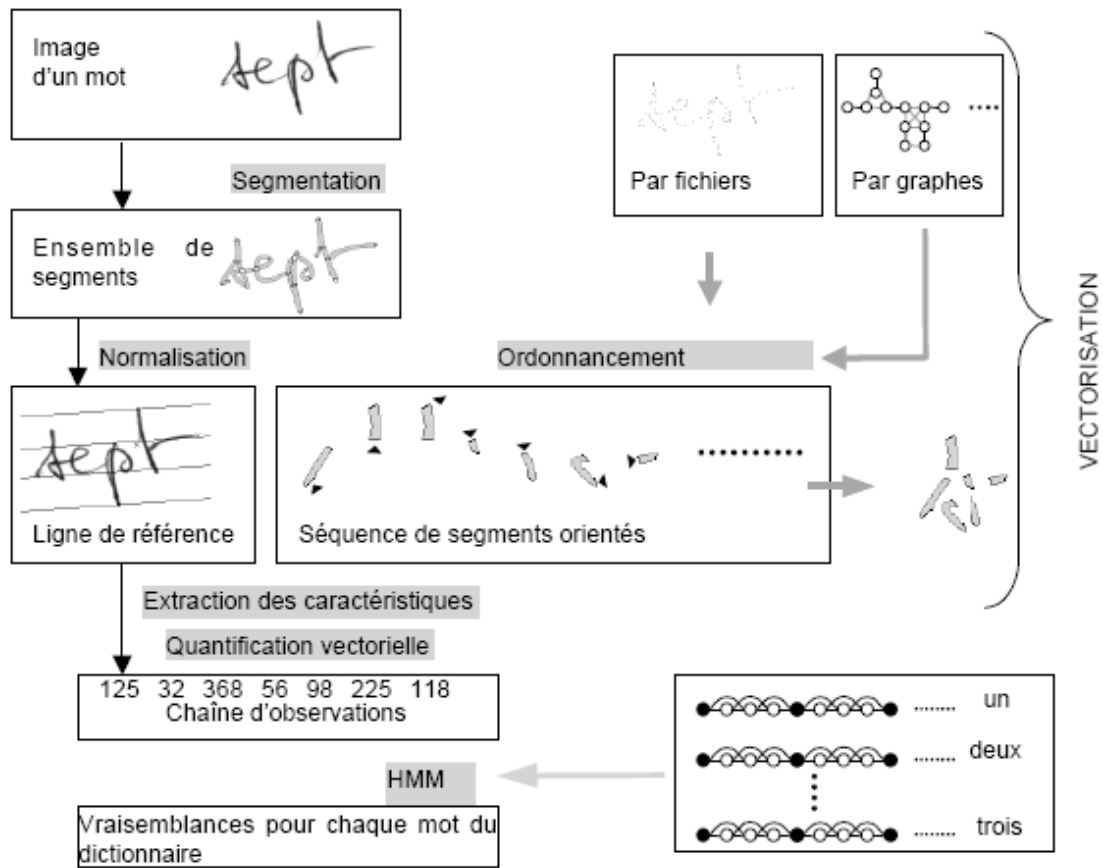


Figure 4.6: Procédure d'utilisation d'un MMC pour la reconnaissance manuscrite [44]

Les mots sont modélisés par la concaténation des MMC des lettres constituant le mot. Exemple Figure 4.7 :

- le MMC du mot « Eté », il est constitué du MMC de la lettre « E » puis « t » et « é »

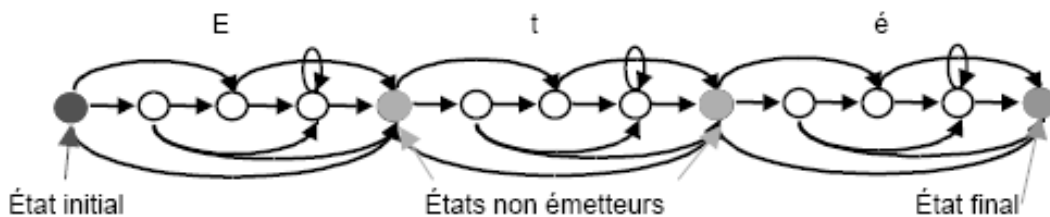
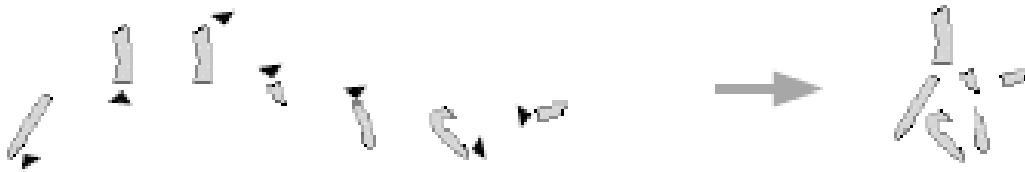


Figure 4.7 : MMC du mot « Eté » [44]

Dans ce genre d'application, le MMC à une topologie Gauche-Droite. En effet, chaque segment représentant un état, il y a une continuité dans la suite des segments ce qui se traduit par un enchaînement logique des différents états. Si le modèle était de type ergodique, cela signifierait que les segments n'ont pas d'ordre précis.

Reprenons l'exemple, la lettre « S » est la continuité des segments suivants :



Il n'est pas naturel d'écrire en inversant les différents segments :



Par contre, un segment pourrait ne pas être reconnu, (lever du stylo, erreur de vectorisation), grâce à cette topologie la reconnaissance n'en sera pas perturbée.



Notant que les MMCs 1D ont été utilisés aussi bien pour la modélisation des variations de l'écriture que pour la résolution du problème de la segmentation en graphèmes. Ces modèles permettent, moyennant un choix judicieux des paramètres, de modéliser correctement l'écriture arabe en atteignant des scores encourageants à l'instar d'autres écritures (en l'occurrence latine) : les MMCs sont transparents à l'alphabet considéré [6].

On peut ainsi distinguer deux extensions des MMC au cas 2D [4, 46]:

- les modèles de Markov pseudo-2D ou planaire,
- les champs de Markov.

Le premier modèle est plutôt un modèle deux fois 1D (MMC simple) alors que le second est intrinsèquement bidimensionnel.

8. Conclusion

Nous avons présenté dans ce chapitre les modèles de Markov cachés qui sont des méthodes stochastiques largement utilisées dans le domaine de la reconnaissance de l'écriture.

Comparés à d'autres approches de reconnaissance, les MMCs se distinguent par leur capacité de modéliser efficacement différentes sources de connaissance. En effet, d'une part ils offrent une intégration cohérente de différents niveaux de modélisation (morphologique, lexical et syntaxique) et d'autre part, il existe des algorithmes puissants pour la reconnaissance et l'apprentissage.

Dans le chapitre suivant, nous allons utiliser les MMC pour développer notre système de reconnaissance.

Chapitre5

Contribution à La reconnaissance des montants littéraux arabes

Résumé : Dans ce chapitre, nous exposerons notre approche de la reconnaissance des montants littéraux arabes, et les algorithmes proposés durant notre recherche dans ce domaine, ainsi que les expériences effectuées sur une base d'images et les résultats obtenus.

1. Introduction

La reconnaissance de l'écriture est un domaine très vaste tant par ses applications que par ses techniques. Dans ce travail, nous nous sommes intéressés à la reconnaissance de l'écriture manuscrite hors-ligne dans le cas des montants littéraux arabes où le nombre de mots à reconnaître constitue un lexique de taille réduite (48 mots). Le système proposé est omni-scripteurs, et basé sur une approche holistique, et sa modélisation se fait à l'aide des MMCs.

2. La base d'images

Dans le cadre de notre travail pour le développement d'un système de reconnaissance des montants littéraux arabes, où le vocabulaire contient 48 mots (voir le tableau 5.1). Nous avons conçu un formulaire pré-imprimé constitué de trois pages dont la page n° 1 contient un entête qui est composé du nom et prénom de scripteur, sa profession, son âge et les 14 premiers mots ; les deux autres pages n° 2 et n° 3 contiennent respectivement les mots de 15 jusqu'à 31 et de 32 jusqu'à 48. 11 scripteurs ont participé à ce travail, et chacun d'eux écrit le mot 4 fois, ce qui nous donne 44 occurrences pour chaque mot ; et 2112 mots arabes au total.

La plupart des travaux de recherche se fondent sur des images de mots déjà segmentés [47]. Dans notre cas, les images de mots sont segmentées manuellement pour mieux familiariser avec les principales caractéristiques morphologiques de l'Arabe et pour prendre en considération toutes leurs variations.

N°	Mot	N°	Mot	N°	Mot	N°	Mot
01	احد	13	عشرون	25	أربعمائة	37	مليوناً
02	اثنان	14	ثلاثون	26	خمسمائة	38	ملايين
03	اثنا	15	أربعون	27	ستمائة	39	مليار
04	ثلاثة	16	خمسون	28	سبعمائة	40	مليارات
05	أربعة	17	ستون	29	ثمانمائة	41	مليارات
06	خمسة	18	سبعون	30	تسعمائة	42	ملايير
07	سنة	19	ثمانون	31	ألف	43	سنتيم
08	سبعة	20	تسعون	32	ألفان	44	سنتيمات
09	ثمانية	21	مائة	33	ألفا	45	دينار
10	تسعة	22	مائتان	34	ألاف	46	دنانير
11	عشرة	23	مائتا	35	مليون	47	و
12	عشر	24	ثلاثمائة	36	مليونان	48	جزائري

Tableau 5.1 : le vocabulaire des montants littéraux arabes

Variations des dimensions	Inclinaisons différentes
سبعون سبعون سبعون	ثمانون احدى اثنا ملياران
Aspects multi scripteurs	Ecritures discontinues
ألاف ألف ألف احد احد احد خمسة خمسة خمسة	عشرون خمسة استحيات دنائير
Connexion de pseudo mots	Ecritures rectifiées
سبعون عشرون مئتا ثلاثون مليونك	ثلاثمائة ثلاثت ثلاثمائة عشوة ملايين خمسة
Elongations variables et omission de Hamza	Espaces inter pseudo-mots variables
أربعة أربعة أربعة	ثمانون ثمانون ثمانون

Tableau 5.2 Exemple des mots illustrant différentes difficultés relatives au manuscrit arabe dans notre base.

3. Le système proposé

A travers les différentes étapes effectuées dans notre système, on peut constater que la structure générale d'un système de reconnaissance hors-ligne de mots manuscrits a été conservée (voir figure 5.6). Dans un premier temps une phase d'acquisition est réalisée afin de numériser l'information manuscrite à reconnaître, suivie d'un ensemble d'opérations de prétraitements permettant d'une part de réduire au maximum le volume de données à manipuler, et d'autre part d'éliminer autant que possible la variabilité, liée à l'écriture. Une opération de transformation de l'image bidimensionnelle en signal monodimensionnel s'est précédée l'étape d'extraction des caractéristiques. Cette étape permet de convertir les données réelles constituant le signal (l'image en une seule dimension) aux coefficients (caractéristiques) pertinents à l'aide des techniques de traitement du signal. Un modèle de Markov caché (MMC) gauche droite est utilisé pour modéliser chaque mot en se basant sur les coefficients extraits du signal.

3.1 Acquisition

Les formulaires ont été scannés à 300 dpi et les images stockées sous le format TIFF (Tagged Image File Format).

3.2 Prétraitements

La littérature montre que les opérations de pré-traitements connues en traitement d'images, ne sont pas toutes appropriées à l'Arabe [34]. Ce qui nécessite de proposer d'autres pré-traitements qui prennent en compte les caractéristiques particulières de l'écriture arabe.

Une fois l'image est numérisée, une série de prétraitements est appliquée. Nous avons utilisés des opérations relativement standard telles que la binarisation, filtrage, lissage...etc. et nous avons aussi proposés d'autres opérations. Ces prétraitements ont pour but la réduction du volume de données à manipuler d'une part et de diminuer la variabilité intra-classe et de l'augmenter inter-classe d'autre part.

3.2.1 Prétraitements classiques

3.2.1.1 Binarisation, filtrage

La plupart des systèmes de reconnaissance utilisent des images en noir et blanc. L'écriture manuscrite étant souvent noire sur fond blanc, les systèmes actuels traitent généralement les images sans utiliser les informations de niveaux de gris ou de couleurs qui sont pourtant fournies par le procédé de numérisation [46]. Les opérations de filtrage de l'image permettent de réduire les bruits d'acquisition [22].

3.2.1.2 Suppression des espaces entourant le mot

Cette opération consiste à supprimer l'espace en droite, gauche, haut et bas du mot ; c'est-à-dire de rendre l'hauteur de la boîte englobante le mot traité est égale à l'hauteur du mot lui-même afin de ne traiter que l'espace géométrique occupé par le mot (voir figure 5.1).

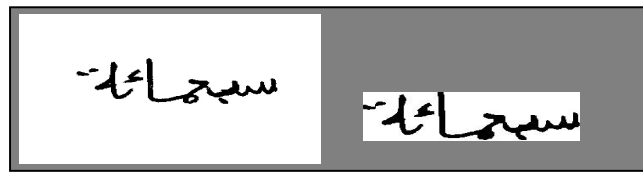


Figure 5.1 : Suppression des espaces entourant le mot «سبعمائة».

3.2.2 Prétraitements proposés

3.2.2.1 Suppression des allongements horizontaux

Les allongements (ligatures) horizontaux et verticaux permettent de mettre en évidence les propriétés structurelles des caractères et des mots. Mais leur principal inconvénient est qu'elles ont des longueurs variables ce qui implique une variabilité intra-mots (voir figure 5.2 a). Pour réduire cette variabilité, nous avons supprimé les colonnes successives redondantes pour ne garder qu'une seule (voir figure 5.2 b). Nous ne supprimons que les allongements horizontaux et nous gardons les allongements verticaux, pour prendre en compte la présence des hampes, des jambages et des points diacritiques.

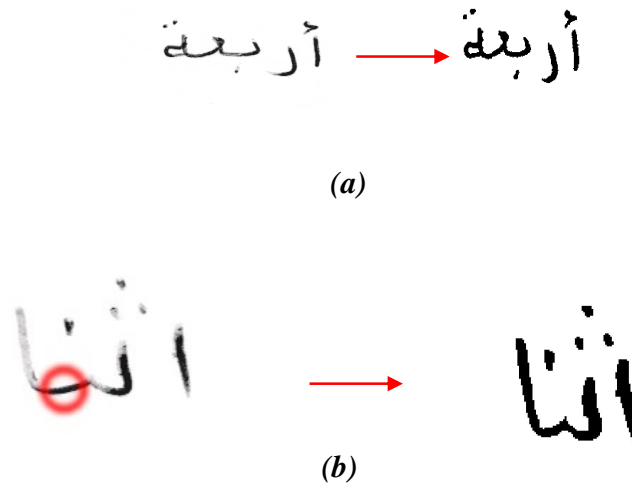


Figure 5.2: (a) Elongations variables du mot « أربعة » (b) : l'image du mot « اثنا » avant et après la suppression de l'allongement horizontal.

3.2.2.2 Suppression des espaces verticaux

En fait, parmi les difficultés rencontrées lors de la reconnaissance de l'écriture arabe manuscrite, on peut citer le problème des positions erronées des points diacritiques au dessus et au dessous des caractères (voir figure 5.3). Pour résoudre ce problème, nous avons supprimé les espaces verticaux.

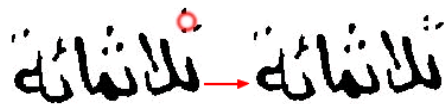


Figure 5.3 : l'image du mot « ثلاثمائة » avant et après la suppression d'un espace vertical

3.2.2.3 Suppression des espaces horizontaux (inter sous-mots).

En plus du problème des positions erronées des points diacritiques ; on note également les espaces inter sous-mots généralement très variables qui rendent la reconnaissance très difficile.

Pour résoudre ce type de problème on a opté pour la suppression de ces espaces rendant ainsi la variabilité plus restreinte (voir figure 5.4).

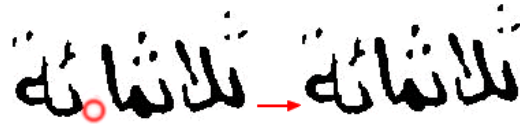


Figure 5.4 : l'image du mot « ثلاثمائة » avant et après la suppression des espaces horizontaux

3.3 Extraction des caractéristiques

L'application des MMCs à la reconnaissance de l'écriture se ramène généralement à la transformation de la forme en primitives judicieusement choisies, traduisant les formes initiales en un signal unidimensionnel composé d'une suite d'observations destinée à être traitée par un modèle de Markov. Les différents travaux réalisés reposent pour une grande part sur l'expérience accumulée dans le domaine de la reconnaissance de la parole où les MMCs sont fréquemment utilisés [34]. Le principe est d'assimiler la dimension horizontale de l'écriture au temps et de calculer des vecteurs de primitives dans une fenêtre parcourant cet axe [46].

Contrairement aux méthodes d'extraction des caractéristiques classiques, la nouvelle méthode proposée permet de calculer les caractéristiques à l'aide des techniques de traitement de signal au lieu de les extraire directement de l'image binaire.

Pour se faire, on procède par le passage d'une représentation bidimensionnelle en une représentation monodimensionnelle de l'image, et ce afin de tirer profit du progrès constaté dans les domaines de la reconnaissance automatique d'objets représentés sous formes de signaux tel que la reconnaissance automatique de la parole.

3.3.1 Processus de conversion de l'image 2D en 1D

Naturellement, l'image a une représentation graphique bidimensionnelle, donc chaque élément de cette représentation (pixel) est identifié par ces coordonnées (i,j) (i sur l'axe d'abscisses, et j sur l'axe d'ordonnés) et sa valeur binaire (0 est l'équivalent de la couleur noire et 1 pour la couleur blanche).

Le processus de conversion consiste à balayer tout l'espace colonne par colonne de droite à gauche et pixel par pixel de haut en bas où chaque pixel est remplacé par deux valeurs. Ces deux valeurs peuvent être négatives si le pixel est blanc, ou positives si le pixel est noir. Et elles sont calculées de la manière suivante :

- en cas d'un pixel noir :

$$\left\{ \begin{array}{l} \text{PX}_{i,j} = i / (\text{ha} + \text{lo}) \\ \text{PY}_{i,j} = j / (\text{ha} + \text{lo}) \end{array} \right. \quad \begin{array}{l} \text{ha: hauteur de l'image, lo: longueur de l'image} \\ \text{(5.1.a)} \\ \text{(5.1.b)} \end{array}$$

- en cas d'un pixel blanc :

$$\left\{ \begin{array}{l} \text{PX}_{i,j} = -i / (\text{ha} + \text{lo}) \\ \text{PY}_{i,j} = -j / (\text{ha} + \text{lo}) \end{array} \right. \quad \begin{array}{l} \text{ha: hauteur de l'image, lo: longueur de l'image} \\ \text{(5.2.a)} \\ \text{(5.2.b)} \end{array}$$

La représentation monodimensionnelle (signal) est construite au fur et à mesure de balayage de l'image donnée :

$$\text{Img}_{1D} = [\text{PX}_{1,1} \text{PY}_{1,1} \text{PX}_{2,1} \text{PY}_{2,1} \text{PX}_{3,1} \text{PY}_{3,1} \dots \text{PX}_{i,j} \text{PY}_{i,j} \dots \text{PX}_{\text{ha}-1,\text{lo}} \text{PY}_{\text{ha}-1,\text{lo}} \text{PX}_{\text{ha},\text{lo}} \text{PY}_{\text{ha},\text{lo}}] \quad (5.3)$$

3.3.2 L'extraction des primitives pour le traitement de la parole [46]

Le processus d'extraction des primitives dans un signal de parole répond aux trois objectifs suivants :

– extraire les caractéristiques pertinentes du signal,

- réduire l'influence du bruit,
- décorrélérer les données,
- diminuer le volume des données.

Ces caractéristiques sont en grande partie dans le domaine fréquentiel, c'est pourquoi une analyse spectrale est nécessaire. Si on garde autant de coefficients que d'échantillons

Dans la fenêtre, ces représentations sont équivalentes, et on peut passer de l'une à l'autre par des transformations simples. L'intérêt est de ne garder que les coefficients les plus discriminants, ces traitements ne sont en pratique pas équivalents. Le signal de parole n'étant pas stationnaire, on réalise une analyse fenêtrée du signal (analyse à court terme).

On détermine les vecteurs de primitives en découpant le signal en fenêtres avant d'extraire les coefficients sur chacune d'elles. En général, on utilise des fenêtres de Hamming de 20 à 30 ms toutes les 10ms de signal.

3.3.2.1 Les bancs de filtres

Une première approche d'extraction de caractéristiques dans un signal de parole consiste à réaliser un banc de filtres pour extraire les coefficients de Fourier. L'analyse de Fourier à court-terme ainsi effectuée permet d'estimer la densité spectrale de puissance à chaque instant.

3.3.2.2 Le codage linéaire prédictif

Une autre méthode simple et efficace pour déterminer les caractéristiques d'un signal comme un signal de parole est le codage linéaire prédictif (LPC en anglais ou modélisation auto-régressive AR) où l'on cherche à estimer la valeur d'un signal x à un instant n en fonction de sa valeur aux instants précédents. Une prédiction d'ordre p détermine les p coefficients a_k tels que le signal

$$\tilde{x}[n] = \sum_{k=1}^p a_k x[n - k], \quad (5.4)$$

soit le plus proche possible de x pour un certain critère (comme l'écart quadratique moyen).

Plusieurs algorithmes, notamment fondés sur la covariance ou l'autocorrélation du signal, existent pour déterminer ces coefficients.

On peut utiliser une échelle de fréquence non linéaire adaptée à la perception de l'oreille humaine, on calcule alors la prédiction linéaire perceptuelle (PLP).

3.3.2.3 Le cepstre

Le calcul du cepstre réalise une analyse fréquentielle du signal avant de repasser dans le domaine temporel : Le principe est de prendre le logarithme du module de la transformée de Fourier du signal (le spectre) et d'en faire la transformée de Fourier inverse (d'où le nom cepstre). L'intérêt de ce traitement est de déconvoluer la fonction de transfert du conduit vocal de l'excitation, qui porte l'information utile. Les coefficients du cepstre sont donc définis par

$$c[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |X(e^{j\omega})| e^{j\omega n} d\omega, \quad (5.5)$$

Où $X(e^{j\omega})$ est la transformée de Fourier du signal de parole. La variable n , homogène à un temps, peut être de la même façon que le cepstre désignée par quéfrenc.

Il est plus efficace d'associer les coefficients cepstraux à une échelle perceptuelle des fréquences, la plus courante aujourd'hui étant l'échelle Mel. Les fréquences de l'échelle Mel sont déduites de la fréquence f par la relation

$$B(f) = 1125 \ln(1 + f/700). \quad (5.6)$$

Ces coefficients sont les *Mel-Frequency Cepstral Coefficients (MFCC)*.

Les vecteurs de primitives pour la reconnaissance de la parole sont typiquement composés des coefficients c_k du traitement MFCC jusqu'au 13e ordre, le coefficient c_0 étant généralement remplacé par l'énergie de la trame de signal.

Afin de rendre mieux compte de la dynamique du signal de parole, on utilise généralement aussi les dérivées des coefficients MFCC. On calcule ces dérivées sur les deux ou quatre échantillons autour de l'échantillon courant, soit

$$\Delta c_k = c_{k+1} - c_{k-1}, \quad (5.7.a)$$

Ou

$$\Delta c_k = c_{k+2} - c_{k-2}, \quad (5.7.b)$$

Ou bien encore

$$\Delta c_k = 2c_{k+2} + c_{k+1} - c_{k-1} - 2c_{k-2}. \quad (5.7.c)$$

On utilise ces coefficients aussi jusqu'au 13^e ordre, et on ajoute enfin la notion d'accélération, elle aussi jusqu'au 13^e ordre :

$$\Delta\Delta c_k = \Delta c_{k+1} - \Delta c_{k-1}. \quad (5.8)$$

On obtient finalement un vecteur de primitives de 39 composantes toutes les 10 ms. La plupart des systèmes aux performances du niveau de celles de l'état de l'art utilisent ce vecteur de primitives dont la pertinence est donc déjà largement démontrée.

3.3.3 Extraction des caractéristiques dans notre système

Nous avons utilisé les 13 premiers coefficients inclus le coefficient c_0 ; et aussi le logarithme de l'énergie, ce qui nous donne 14 coefficients auxquels nous avons ajouté leurs dérivées premières et secondes pour prendre en compte la dynamique des observations. On obtient donc 42 coefficients par vecteur.

Le résultat de cette étape est une séquence de vecteurs des caractéristiques de l'image du mots considéré représentée par une matrice de taille : $42 \times \text{nombre_vecteurs}$, où nombre_vecteurs dépend de la taille du mot en pixel (hauteur x longueur).

3.4 Apprentissage

Nous avons élaboré un MMC = $\{\pi, A, \mu_i, \Sigma_i \mid i = 1, \dots, N\}$ du type strictement gauche droite pour chaque mot du vocabulaire (π : La distribution d'état initial, A : matrice de transition, μ_i : vecteur moyen, Σ_i : matrice de covariance) avec les éléments de la matrice de transitions $A(N \times N)$ $a_{ij} \neq 0$ pour $j=i$ et $j=i+1$.

Après plusieurs expériences, nous avons fixé le nombre d'états à $N=17$ et comme exception, le modèle du mot « و » a 4 états uniquement (le nombre de vecteurs des caractéristiques pour ce mot souvent est inférieur à 10).

L'observation émise par chacun des états est modélisée par une densité de probabilité mono-gaussienne.

Nous avons utilisé le critère de Viterbi (décrit dans le chapitre4) à la base de l'algorithme d'apprentissage pour le raison suivant :

- l'algorithme d'apprentissage utilisant le meilleur chemin est suffisant puisque le taux de reconnaissance obtenu par cet algorithme et celui obtenu par Baum-Welch diffère de moins de 0,5 % [42]. Et en plus, il peut réduire la complexité du système.

La figure 5.1 montre que les probabilités convergent vers une valeur limitée, après 10 itérations.

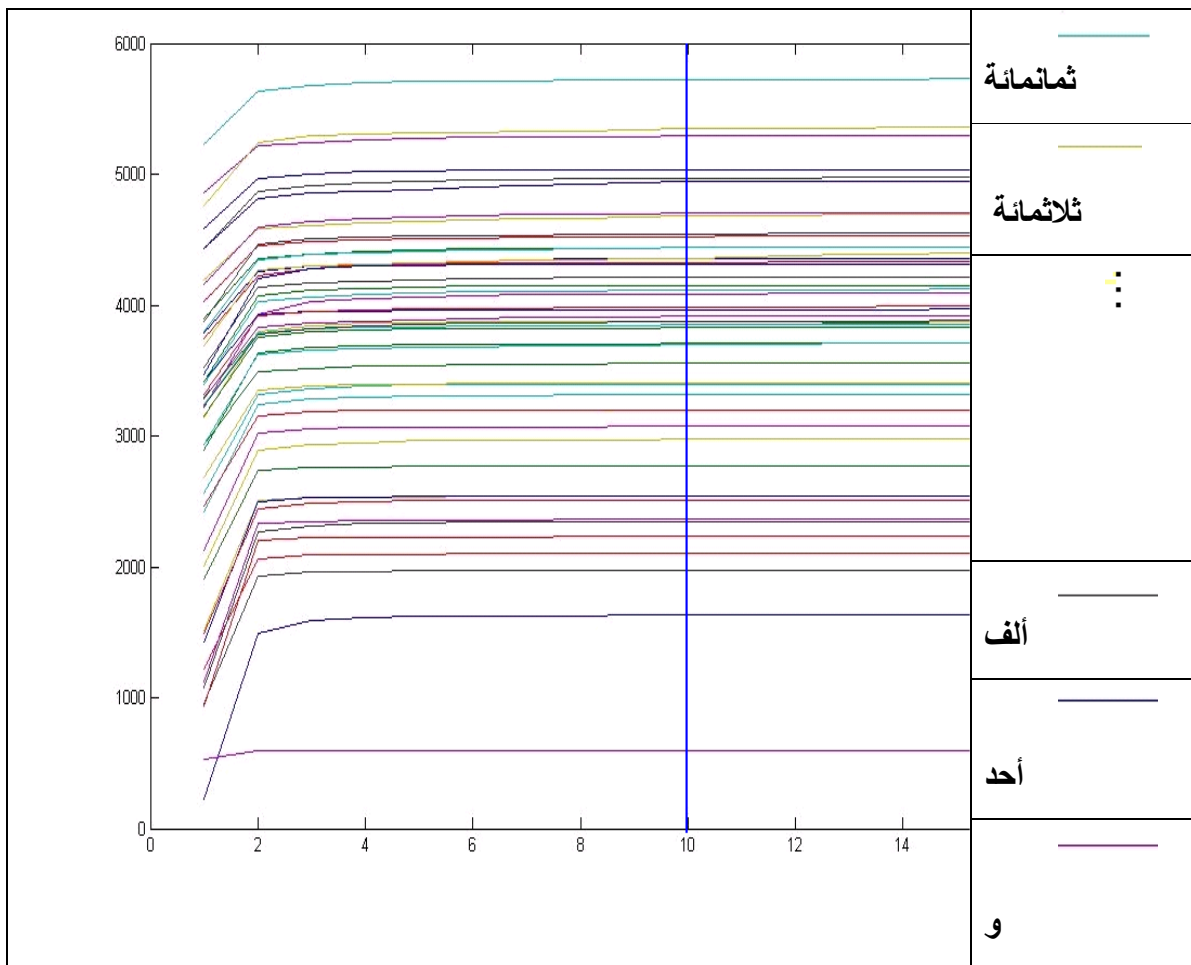


Figure 5.5 : les probabilités lors des itérations d'apprentissage pour chaque mot

3.5 La reconnaissance

La reconnaissance se fait aussi par l'algorithme de Viterbi, où on cherche le meilleur chemin correspond à la séquence d'observations (mot à reconnaître) au sens du maximum de logarithme de vraisemblance.

Nous avons préféré l'utilisation du logarithme de vraisemblance, pour [46]:

- éviter de manipuler des probabilités qui deviennent rapidement inférieures à la précision de la machine,
- calculer des additions plutôt que des multiplications.

L'image du mot à reconnaître est d'abord préparée en utilisant le module de prétraitement, puis convertie d'une image (2D) en un signal (1D). Le signal obtenu est ensuite paramétrisé par la méthode MFCC pour extraire les vecteurs des caractéristiques. Le classifieur markovien calcul à partir de ces vecteurs le logarithme de vraisemblance de mot à reconnaître par rapport à tous les modèles de mots de référence, et le mot reconnu est celui qui a la valeur maximale.

3.5.1 Stratégies de rejet

Dans les applications pratiques [46], on utilise une politique de rejet qui évite de prendre une décision quand celle-ci serait trop incertaine. Plusieurs stratégies peuvent être envisagées :

- une stratégie de rejet par seuil absolu où seules les probabilités supérieures à un seuil sont retenues,
- une stratégie par seuil relatif où la décision n'est prise que si la différence entre les deux plus grandes probabilités est supérieure à un seuil.

Le rejet relatif est nettement plus efficace et permet d'obtenir un taux d'erreurs inférieur à 1% pour un rejet de 10% des images d'entrée.

Dans notre système, nous avons utilisé le rejet relatif où la décision n'est prise que si le rapport entre les deux plus grandes probabilités ($\text{prob1} / \text{prob2}$) est supérieure à 1.005.

La figure 5.6 illustre l'architecture générale du système proposé.

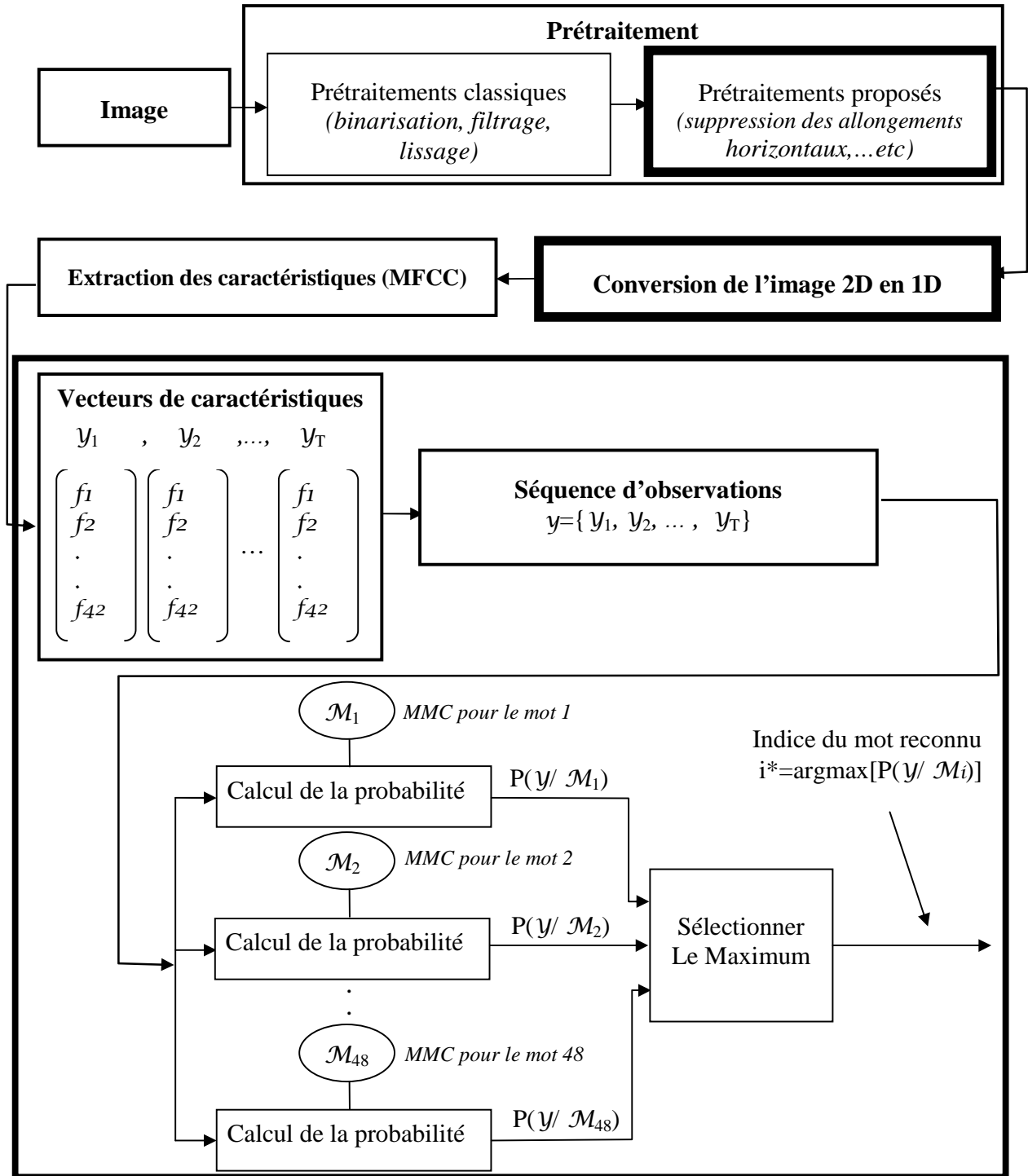


Figure 5.6 : Architecture du système proposé de reconnaissance de mots arabes

4. Les algorithmes

```

=pre-tr_supp_allongement_horz(img0) ;
    Pour chaque colonne col                                % de 1 à taille_image(img0,ind_col)
        Si col ≡ col+1                                    % les deux colonnes sont identiques
            Alors
                Supprimer(col+1) ;
            fin si
        fin pour                                        % chaque colonne
fin pre-tr_supp_allongement_horz;

```

Algorithme 5.1 : l'algorithme de suppression des allongements horizontaux

```

=pre-tr_supp_esp_vert(img0) ;
    Pour chaque ligne lign                                % 1 à taille_image(img0,ind_lign)
        Si lign ≡ espace                                  % tous les pixels de lign sont blancs
            Alors Si lign+1 ≡ espace                    % tous les pixels de lign+1 sont blancs
                Alors
                    Supprimer(lign+1) ;
                fin si
            fin si
        fin pour                                        % chaque ligne
fin pre-tr_supp_esp_vert;

```

Algorithme 5.2 : l'algorithme de suppression des espaces verticaux

```

=pre-tr_supp_esp_horz(img0) ;
    Pour chaque colonne col                                % 1 à taille_image(img0,ind_col)
        Si col ≡ espace                                  % tous les pixels de col sont blancs
            Alors Si col+1 ≡ espace                    % tous les pixels de col+1 sont blancs
                Alors
                    Supprimer(col+1) ;
                fin si
            fin si
        fin pour                                        % chaque colonne
fin pre-tr_supp_esp_horz ;

```

Algorithme 5.3 : l'algorithme de suppression des espaces horizontaux


```

modeles=apprentissage(caracters, N); % caracters : caractéristiques
    Pour chaque mot % de 1 à 48
        A=initializ_A(N(mot)); % A initiale
        Pour chaque occurrence occ % de 1 à 36
            [MI(occ),SIGMA(occ)]= initializ_MI_SIGMA(caracters(mot ,occ),N(mot ));
            fin pour % chaque occurrence occ
            MI=moyen(MI(1 : nb_occurrences )); % MI initiale
            SIGMA=moyen(SIGMA(1 : nb_occurrences )); % SIGMA initiale
            modele_initial (A, MI , SIGMA );
            seq=les nb_occurrences du mot ; % nb_occurrences =36
            modele(A, MI , SIGMA ) = Reestimation(seq,modele_initial(A, MI , SIGMA ));
            % par Viterbi
            Tantque nb_iter<=10 % 10 itérations
                modele(A, MI , SIGMA ) = Reestimation( seq,modele(A, MI , SIGMA ));
                % par Viterbi
            fin tantque
            modeles(mot)=modele(A, MI , SIGMA );
        fin pour % chaque mot
fin apprentissage ;

```

Algorithme 5.4 : l'algorithme d'apprentissage du système proposé

```

mot-reconnu =test(img-mot-à-reconnaître, modeles) ;
    img-mot-à-reconnaître =pré-traitements(img-mot-à-reconnaître) ;
    signal-mot-reconnaître=conversion-2Den1D(img-mot-à-reconnaître) ;
    caractères-mot-à-reconnaître=extract-caractères(signal-mot-reconnaître) ;
    Pour chaque modele mod % 1 à 48
        prob(mod) = viterbi_log (caractères-mot-à-reconnaître, modeles(mod));
    fin pour % chaque modele
    mot-reconnu=trier-descendant(prob) ; % mot-reconnu où prob est maximum ;
fin test ;

```

Algorithme 5.5 : l'algorithme de reconnaissance du système proposé

5. Interprétation des résultats

Les modèles ont été entraînés sur une partie de la base (1728 mots) et les tests ont été effectués sur le reste de la base (384 mots).

Le tableau 5.3 présente les résultats du système proposé sans et avec combinaison de trois prétraitements proposés. La valeur du taux de reconnaissance (resp. taux d'erreur), pour le choix 1 (choix n = les n meilleurs candidats inclus le mot approprié.), croît de 73.44% à 87.24% (resp. décroît de 26.56% à 12.76%). Ce qui montre que les prétraitements proposés produisent une amélioration notable.

	Prétraitement1 <i>Suppression des espaces horizontaux</i>	Prétraitement2 <i>Suppression des espaces verticaux</i>	Prétraitement3 <i>suppression des allongements horizontaux</i>	taux
choix1	✘	✘	✘	73.44%
	✔	✘	✘	77.34%
	✘	✔	✘	78.38%
	✘	✘	✔	70.83%
	✔	✔	✘	79.68%
	✔	✘	✔	75.78%
	✘	✔	✔	83.07%
	✔	✔	✔	87.24%

✔ Le prétraitement correspond est utilisé

✘ Le prétraitement correspond est non utilisé

Tableau 5.3 : les résultats avant et après l'utilisation des prétraitements proposés

Environ 87.24% de mots de la base ont été affecté à une classe correcte (bien reconnus) alors que 7.29% ont été affecté à une classe incorrecte (substitution), avec 5.47% de mots rejetés.

Choix n	Taux
choix 1	87.24%
choix 2	91.14%
choix 3	92.97%
choix 4	93.75%
choix 5	94.53%

Tableau 5.4 : les taux de reconnaissance.

La substitution se produit quand un mot testé est affecté à une autre classe que celle à laquelle il devrait appartenir. Nous avons constaté que ce problème est généralement causé par les caractères penchés (« لا » et « ا ») et dégradation des boucles (« ف » a été reconnu comme étant « د ») par exemple : « ألف » a été reconnu comme étant « أحد » (voir figure 5.7).

Deux échantillons du mot « ثلاثمائة » (voir resp. figure 5.7 et figure 5.8) ont été reconnus comme étant resp. « ثلاثة », « ثلاثون » à cause de l'épaisseur du trait d'écriture et les variations des dimensions.

Le problème de rejet s'est posé avec un échantillon « ألف » (voir figure 5.7) contenant une petite hampe penchée. L'échantillon a été reconnu comme étant « ألف » mais le prob1/ prob2 est inférieur à 1.005. Nous avons aussi rencontré le problème inverse au cours de nos expérimentations.

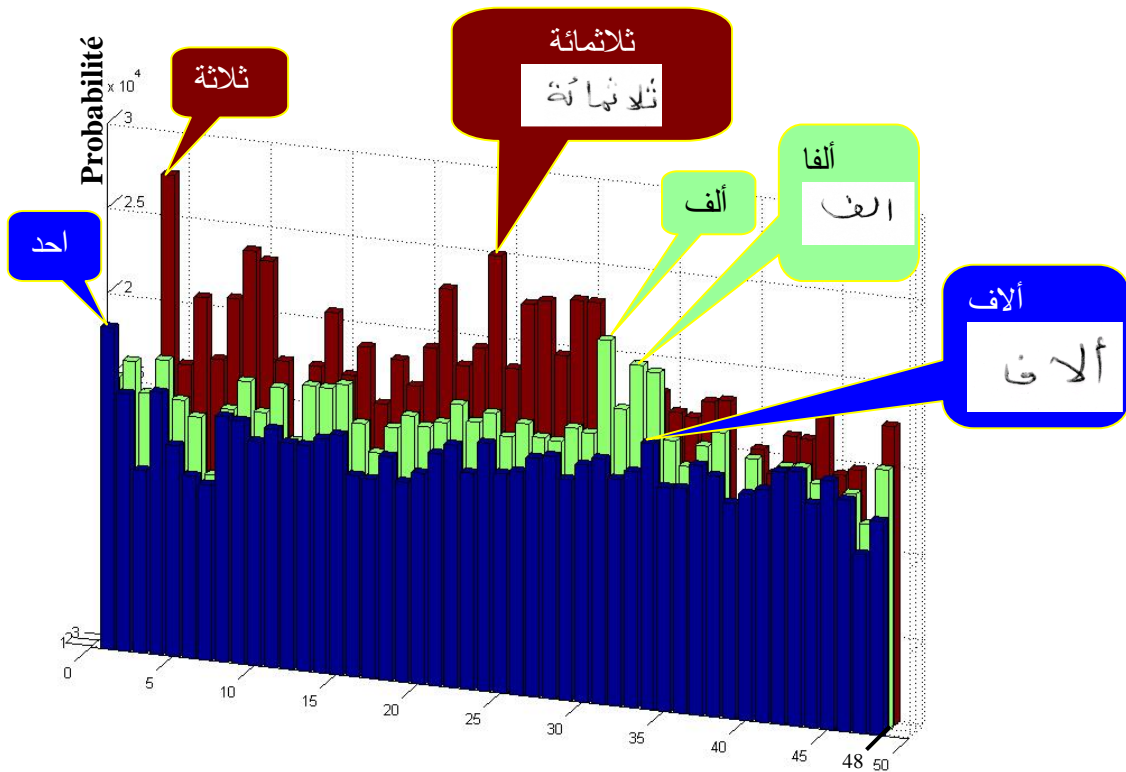


Figure 5.7 : échantillons de mots rejetés ou mal reconnus

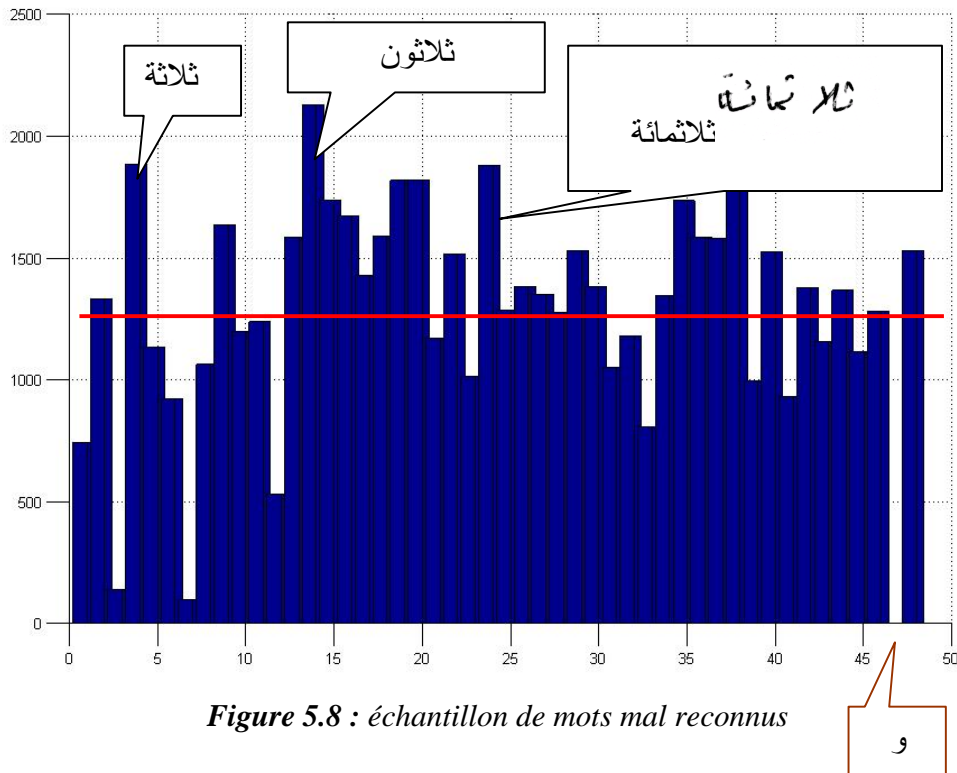


Figure 5.8 : échantillon de mots mal reconnus

La figure 5.9 montre que le mot « و » est souvent très bien reconnu à cause de sa morphologie particulière.

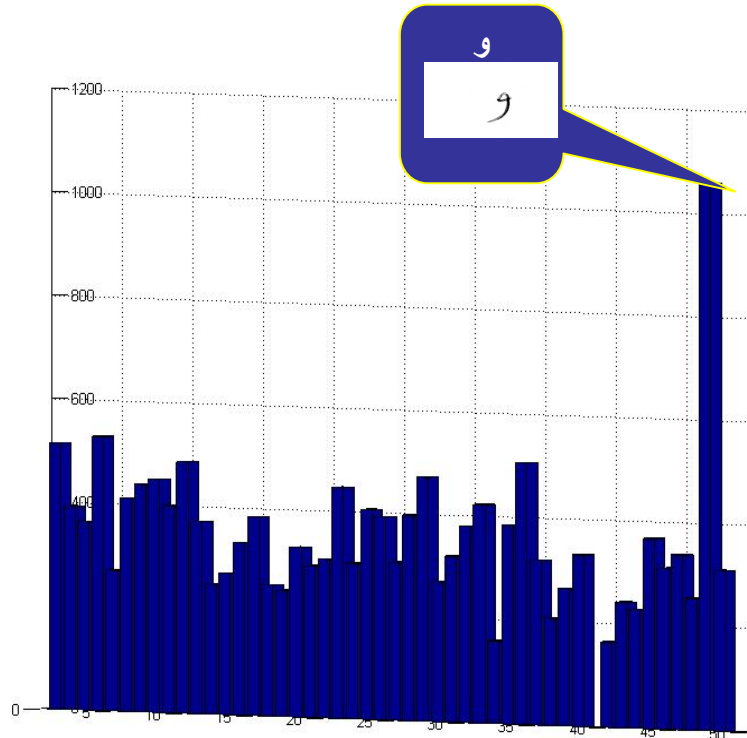


Figure 5.9 : échantillon de mot « و » très bien reconnu

6. Amélioration du système par combinaison des classifieurs

La combinaison de classifieurs est un outil efficace pour avoir une grande performance sans augmenter la complexité des techniques de classification existantes. Elle est convenable pour les applications nécessitant une grande exactitude de classification [9].

Pour être combinés dans un système multiclassifieurs, les classifieurs de base doivent être différents. La création de la diversité entre les deux classifieurs à combiner dans notre système se fait pendant l'étape de conversion de l'image (2D) en un signal (1D) par changement de la direction de parcours de l'image, et la fonction de décision utilisée est la somme des logarithmes des vraisemblances issus de chaque classifieur.

Les résultats de l'évaluation de chaque classifieur à part, et du système après la combinaison des deux classifieurs sont comme suit :

- Sans prétraitements proposés

	Classifieur1 : C1 (parcours de droite à gauche)	Classifieur2 : C2 (parcours de gauche à droite)	Classifieur3: C3 (Combinaison de C1 & C2)
Choix 1	73.44	78.64	79.95
Choix 2	83.07	85.94	89.32
Choix 3	87.50	88.80	91.41
Choix 4	90.62	90.88	92.19
Choix 5	91.14	91.93	92.71
Rejet classifieur1 = 06.51%, Rejet classifieur2 = 04.16%, Rejet classifieur3 = 04.94%,			

- Avec prétraitements proposés

	Classifieur1 : C1 (parcours de droite à gauche)	Classifieur2 : C2 (parcours de gauche à droite)	Classifieur3: C3 (Combinaison C1 & C2)
Choix 1	87.24	87.24	90.10
Choix 2	91.14	90.88	93.22
Choix 3	92.97	93.49	94.79
Choix 4	93.75	94.53	95.31
Choix 5	94.53	94.79	95.31
Rejet classifieur1 = 05.46%, Rejet classifieur2 = 04.68%, Rejet classifieur3 = 04.16%.			

La combinaison des deux classifieurs croît le taux de reconnaissance de 87.24% (resp. 73.44%) pour classifieur1, et 87.24% (resp. 78.64%) pour classifieur2, à 90.10% (resp. 79.95%) pour le classifieur combiné dans le cas d'utilisation des prétraitements proposés (resp. sans l'utilisation des prétraitements proposés).

Nous avons constaté à partir des résultats obtenus, que la combinaison des classifieurs peut donner de bons résultats.

Conclusion et Perspectives

1. Conclusion générale

Malgré les efforts et les travaux réalisés dans le domaine de la reconnaissance des chèques arabes, le système idéal n'existe pas. Et un bon système nécessite de faibles taux de rejets et des taux d'erreurs très faibles car le coût des erreurs est extrêmement élevé. Cela dépend d'une grande part des techniques utilisées au cours du processus de développement.

Dans le cadre de notre travail, un système pour la reconnaissance hors-ligne des montants littéraux arabes est développé. Le système repose sur une approche globale, la classification se fait à l'aide des modèles de Markov cachés, et l'extraction d'un ensemble de caractéristiques robustes et indépendantes du langage d'écriture est effectuée en utilisant les techniques de traitement de signal courantes dans le domaine de la reconnaissance de la parole. Ces techniques sont appliquées sur l'image prétraitée après conversion à un signal monodimensionnel.

Le prétraitement des images avant l'extraction des caractéristiques joue un rôle primordial dans la qualité des résultats. Et comme les méthodes de prétraitement couramment utilisées dans la littérature ne prennent pas en compte tous les caractéristiques de l'écriture manuscrite arabe, nous avons proposé quelques techniques de prétraitement.

Notre contribution s'articule autour du quatre points suivants:

- Les prétraitements (suppression des allongements horizontaux, suppression des espaces verticaux et suppression des espaces horizontaux).
- La conversion de l'image 2D en un signal 1D.
- L'exploitation directe des méthodes d'extractions des caractéristiques du signal vocal dans le domaine de reconnaissance de l'écriture.
- Combinaison de classifieurs markoviens différents dans la direction de parcours des images durant l'opération de conversion de l'image (2D) en un signal (1D).

Grâce à l'introduction de ces propositions, notre système donne des résultats très encourageants, dans ce qui concerne les performances et la complexité.

2. Perspectives

Malgré les résultats obtenus, plusieurs tâches restent encore à achever aux différents niveaux :

- Le prétraitement : tel que l'adaptation des lignes de base pour chaque élément d'un mot composé, la correction des caractères penchés,...etc.
- La proposition et l'examen de nouvelles caractéristiques en marquant les squelettes, les contours des mots,...etc (lors de la conversion de l'image en signal).
- Plusieurs paramètres dans l'étape de modélisation avec les MMCs sont ajustables telles que la fréquence d'échantillonnage, la taille de la fenêtre de Hamming,...etc. et l'utilisation des algorithmes génétiques pour la sélection des nombres d'états.
- Une perspective majeure consiste à établir la phase de post-traitement en combinant plusieurs classifieurs (plus que deux).
- Adapter notre système pour d'autres champs d'application tels que la reconnaissance de visages.

Références

[1] M. Kunt, G. Coray, G. Granlund J.P. Haton, R. Ingold, M. Kocher, **Reconnaissance des formes et analyse des scènes**, Collection technique et scientifique des télécommunications du CENT, ver. 3, 2000.

[2] C. Gagne, «**Algorithmes Evolutionnaires Appliqués A La Reconnaissance Des Formes Et à La Conception Optique**», Thèse de doctorat, Université Laval, 2005.

[3] C. Cariou, J.M. Ogier, S. Adam, J. Gardes, R. Mullot, Y. Lecourtier, «**Reconnaissance de Formes en Multi-Orientation et Multi-Echelle sur Documents Techniques**», Dix-septième colloque GRETSI, pp.13-17, Vannes, Septembre 1999.

[4] G. Tremblay, «**Optimisation d'ensemble de classifieurs non paramétriques avec apprentissage par représentation partielle de l'information**», Thèse de doctorat, Ecole de technologie supérieure, Université du Québec, 2004.

[5] N. Benahmed, «**Optimisation de Réseaux de Neurones Pour la Reconnaissance des Chiffres Manuscrits Isolés, Sélection et Pondération des Primitives par Algorithmes Génétiques**», Thèse pour l'obtention de la Maîtrise en Génie de la Production Automatisée, Montréal, Mars 2002.

[6] C. Faure, «**Interaction Homme-Machine**», Document pour l'école jeunes chercheurs CNRS, Disponible sur le lien web : <http://www.tsi.enst.fr/~cfaure/intro/Intro1.html>, Juillet 1997.

[7] R.O. Duda, P.E. Hart, D.G. Stork, **Pattern Classification**, Second edition. New York: Wiley-Interscience, 2001.

[8] C. Saint-Jean, «**Classification paramétrique robuste partiellement supervisée en reconnaissance des formes** », Thèse de doctorat, Université de la Rochelle, 2001.

[9] I. Gasmi, H. Merouani, L. Souici-Meslati, « **Combinaison de classifieurs** », 3rd International Conference: Sciences of Electronic, Technologies of Information and telecommunications SETIT2005, Tunisie, Mars 2005.

[10] L. Miclet, J.P. Haton, **Méthodes structurelles pour la reconnaissance des formes**, Editeur Eyrolles, 1984

[11] D. Arrivault¹, N. Richard, C. Fernandez-Maloigne, P. Bouyer, «**Collaboration entre approche statistique et structurelle pour la reconnaissance de caractères anciens**», 8^{ème} Colloque International Francophone sur l'Écrit et le Document (CIFED'04), vol.1, pp.197-202, La Rochelle, Juin 2004.

[12] F. Grandidier, R. Sabourin, C.Y. Suen, «**Quelques techniques pour l'amélioration du pouvoir discriminant de primitives discrètes**», Conférence Internationale Francophone sur l'Écrit et le Document CIFED'04, 2004.

[13] N. Ben Amara, «**Utilisation des modèles de Markov cachés planaires en reconnaissance de l'écriture arabe imprimée**», Thèse de doctorat ; Université de Tunis II, 1999.

[14] E. Baudrier, «**Comparaison d'images binaires reposant sur une mesure locale des dissimilarités Application à la classification**», thèse de doctorat, Université de Reims Champagne-Ardenne, 2005.

[15] M. Delalandre, P. Heroux, S. Adam, E. Trupin, J.M. Ogier, «**Une approche statistico-structurelle pour la reconnaissance de symboles exploitant une représentation XML des données** », CIFED'02, 2002.

[16] A. Belaid, G. Saon, «**Utilisation des processus markoviens en reconnaissance de l'écriture**», Traitement du Signal, vol. 14, n°. 2, pp. 161 – 177, 1997.

[17] A. Belaid, «**Reconnaissance automatique de l'écriture et du document** », Pour la science, disponible sur le lien web :<http://webloria.loria.fr/~abelaid/Publications.html>, 2001.

[18] C. Chatelain, «**Extraction de séquences numériques dans des documents manuscrits quelconques**», Thèse de doctorat ; Université de Rouen, 2006.

[19] S. Carbonnel, E. Anquetil, «**Apprentissage automatique d'une distance d'édition dédiée à la reconnaissance de l'écriture manuscrite**», *Conférence Internationale Francophone sur l'Ecrit et le Document CIFED'04*, 2004.

[20] R. Plamondon, S. Srihari, «**On-line and Off-line Handwriting Recognition: A Comprehensive Survey**», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n°. 1, pp. 63-84, 2000.

[21] N.E. Ayat, «**Selection de modele automatique des machines a vecteurs de support: application a la reconnaissance d'images de chiffres manuscrits**», *Thèse de doctorat*, 2004.

[22] J.P. Crettez, G. Lorette, «**Reconnaissance de l'écriture manuscrite** », *Techniques de l'Ingénieur, Dossier : H1358, Date de parution : 02/1998*

[23] C.C. Tappert, C.Y. Suen, T. Wakahara, «**The state of the art in online handwriting recognition**», *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, n°. 8, pp.787-808, 1990.

[24] S. Ouchtati, «**Segmentation et Reconnaissance des Chaînes Numériques Manuscrites**», *Thèse de doctorat, université Annaba*, 2007.

[25] B. Al Badr, S.A. Mahmoud, «**Survey and bibliography of Arabic optical text recognition**», *signal processing of ICDAR'95*, pp.355-359, 1995.

[26] C. Choisy, A. Belaid, «**Apprentissage croisé en reconnaissance analytique de l'écriture manuscrite**», *Presses polytechniques et universitaires romandes, Collection des sciences appliquées de l'INSA de Lyon, In CIFED'00*, 2000.

[27] N. Ben Amara, S. Gazzah, «**Une approche d'identification des fontes arabes**», *CIFED'04*, pp.273-278, La Rochelle, France, Juin 2004.

[28] O. Hachour, «**Reconnaissance hybride des caractères Arabes imprimé**», *JEP-TALN, Fès, Avril 2004*.

- [29] H. Miled, «*Stratégies de reconnaissance de l'écriture semi cursive : application aux mots manuscrits arabes*», Thèse de doctorat, Université de Rouen, 1998.
- [30] S. Chevalier, M. Lemaître, E. Geoffrois, «*Étude de primitives spectrales pour la reconnaissance de caractères manuscrits dans le cadre d'une approche markovienne 2D*», Actes 15ème Congrès Francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle (RFIA'2006), Tours, France (CDROM), 2006.
- [31] http://www.cfo-news.com/A2iA-demeure-a-la-pointe-de-l-innovation-sur-son-marche-historique-des-moyens-de-paiement_a1726.html. Source : <http://www.a2ia.com/> Mardi 27 Mars 2007.
- [32] Y. Al-Ohali, M. Cheriet, C. Suen, «*Databases for recognition of handwritten Arabic cheques*», Pattern Recognition vol. 36 , pp. 111-121, 2003.
- [33] P. Burrow, «*Arabic handwriting recognition*», Master of science thesis, School of Informatics, university of Edinburg, England, 2004.
- [34] N. Ben Amara, A. Belaïd, N. Ellouze, «*Utilisation des modèles markoviens en reconnaissance de l'écriture arabe : Etat de l'art*», 3^{ème} Colloque International francophone sur l'écrit et le document CIFED'00, pp.181-191, Lyon, FRANCE, 2000.
- [35] N. Ben Amara, A. Belaïd, N. Ellouze, «*Modélisation pseudo bidimensionnelle pour la reconnaissance des chaînes de caractères arabes imprimées*». 1^{er} Colloque International francophone sur l'écrit et le document CIFED'98, pp.131-140, Quebec, Canada , 1998.
- [36] S. HAITAAMAR, «*Segmentation de textes en caracteres pour la reconnaissance optique de l'écriture arabe*», Thèse de Magister , Université El-Hadj Lakhdhar, Batna, 2007.
- [37] S. Alma'adeed, C. Higgins, D. Elliman, «*Off-line recognition of handwritten Arabic words using multiple hidden Markov models*», Knowledge-Based Systems, vol. 17, pp.75–79, 2004.
- [38] N. Farah, L. Souici, M. Sellami, «*Classifiers combination and syntax analysis for Arabic literal amount recognition*», Engineering Applications of Artificial Intelligence, vol. 19, pp. 29-39, 2006.

- [39] L. Souici-Meslati, M. Sellami, «**A hybrid approach for arabic literal amounts recognition**», *AJSE, the Arabian Journal for Science and Engineering*, Volume 29, Number 2B, pp. 177-194, October 2004.
- [40] L.R. Rabiner, «**A Tutorial on Hidden Markov Models and Select Application in Speesh Recognition**», *Procceding of the IEEE*, vol. 77, n° 2, pp.257-286, 1989.
- [41] A. El Yacoubi, «**Modélisation markovienne de l'écriture manuscrite, application à la reconnaissance des adresses postales**», *Thèse de doctorat, Université de Rennes 1*, 1996.
- [42] J. Bruno, «**Un outil informatique de gestion de Modèles de Markov Cachés : expérimentations en Reconnaissance Automatique de la Parole**», *Thèse de Doctorat de l'Institut de Recherche en Informatique de Toulouse, l'Université Paul Sabatier de Toulouse III, Spécialité Informatique*, 1995.
- [43] S. Aupetit, N. Monmarché, M. Slimane , «**Utilisation des Modèles de Markov cachés pour la reconnaissance robuste d'images : apprentissage par colonie de fourmis, algorithme génétique et essaim particulaire. Optimisation en traitement du signal et de l'image**», *Hermès-Lavoisier*, 2007.
- [44] J. Dequier, «**Chaînes de Markov et applications**», *Examen probatoire en informatique du CNAM, centre régional RHÔNE-ALPES, centre d'enseignement de GRENOBLE*, 2005.
- [45] A. Hlaoui, «**Reconnaissance de mots isolés arabes par hybridation de réseaux de neurones, et modèles de Markov cachés**», *diplôme d'études approfondies analyse des systèmes et traitements numériques option traitement de signal, école nationale d'ingénieurs de Tunis*, 1999.
- [46] S. Chevalier, «**Reconnaissance d'écriture manuscrite par des techniques markoviennes : une approche bidimensionnelle et générique**», *thèse de doctorat, l'université René Descartes - Paris 5, décembre 2004*.
- [47] R.G. Casey, E. Lecolinet, «**A survey of methods and strategies in character segmentation**», *IEEE transactions on pattern analysis and machine intelligence*, vol. 18, n°. 7, pp.690–706, 1996.

Annexe A

Application du système proposé sur la base IFN/ENIT

1. Base d'image IFN/ENIT

la base de données de référence IFN/ENIT de noms de villes tunisiens, est une base de mots arabes manuscrits dont le développement rentre dans le cadre d'un projet de coopération entre le laboratoire de systèmes et de traitement de signal de l'ENIT (Ecole Nationale d'Ingénieurs de Tunis) et le laboratoire de recherche sur la reconnaissance de l'écriture arabe de l'IFN (Insitute Fur Nachrichtentechnik, Insitute of Communications Technology) de l'université de Braunschweig, Allemagne.

Cette base (ifnenit_v2.0p1e) contient des noms manuscrits de 937 villes/villages tunisiens avec leurs codes postaux. 411 scripteurs ont participé à la constitution de la base en remplissant plus de 2200 formulaires qui ont été scannés à 300 dpi, puis les images de noms de villes ont été extraites (sous formats TIFF et BMP).

2. Expériences et résultats

Les modèles ont été entraînés et testés sur une partie de la base IFN/ENIT : (6823 mots pour l'apprentissage, et 2650 mots pour le test et avec un dictionnaire de taille 59 pour les noms de villes). La valeur du taux de reconnaissance croît de 70.78% (sans l'utilisation des prétraitements proposés) à 85.60% (avec l'utilisation des prétraitements proposés). Les résultats obtenus sont illustrés dans les tableaux suivants :

	Prétraitement1 Suppression des espaces horizontaux	Prétraitement2 Suppression des espaces verticaux	Prétraitement3 suppression des allongements horizontaux	Taux %
Top1	☒	☒	☒	70.78
	☑	☒	☒	75.09
	☒	☑	☒	75.28
	☒	☒	☑	75.55
	☑	☑	☒	79.94
	☑	☒	☑	80.46
	☒	☑	☑	82.63
	☑	☑	☑	85.60

Tableau 1 : les résultats avant et après l'utilisation des prétraitements proposés

Top n= les n meilleurs candidats inclus le mot approprié.	Taux %
Top 1	85.60
Top 2	93.89
Top 5	96.05

Tableau 2 : Les taux de reconnaissance

Annexe B ***Interface et Manuel d'utilisation***

Cette annexe est une présentation de l'interface et du manuel d'utilisation de notre système. Une fois le système est déclenché, vous pouvez effectuer les opérations suivantes :

- Vous avez le choix de consulter les deux bases d'images, à partir du menu « Bases d'images », ou avec le raccourci clavier « Ctrl+M » pour la base d'images des mots des montants littéraux arabes, ou avec « Ctrl+I » pour la base d'images IFN/ENIT (voir figure1).

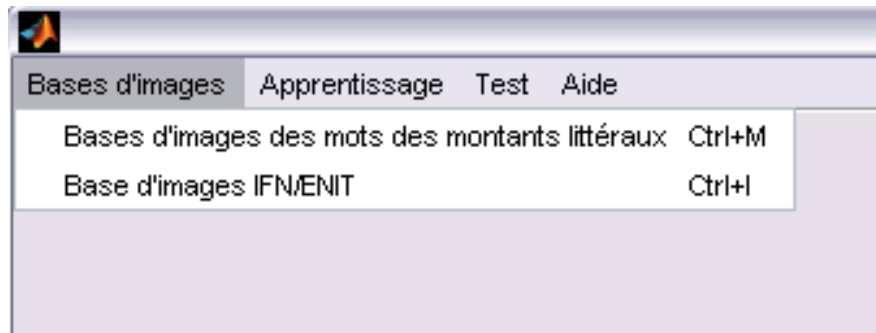


Figure1 : le menu « *Bases d'images* »

Après choisir la base d'images, une nouvelle fenêtre est apparue (voir figure2 pour le choix « Bases d'images des mots des montants littéraux » et figure3 pour le choix « Bases d'images IFN/ENIT »).

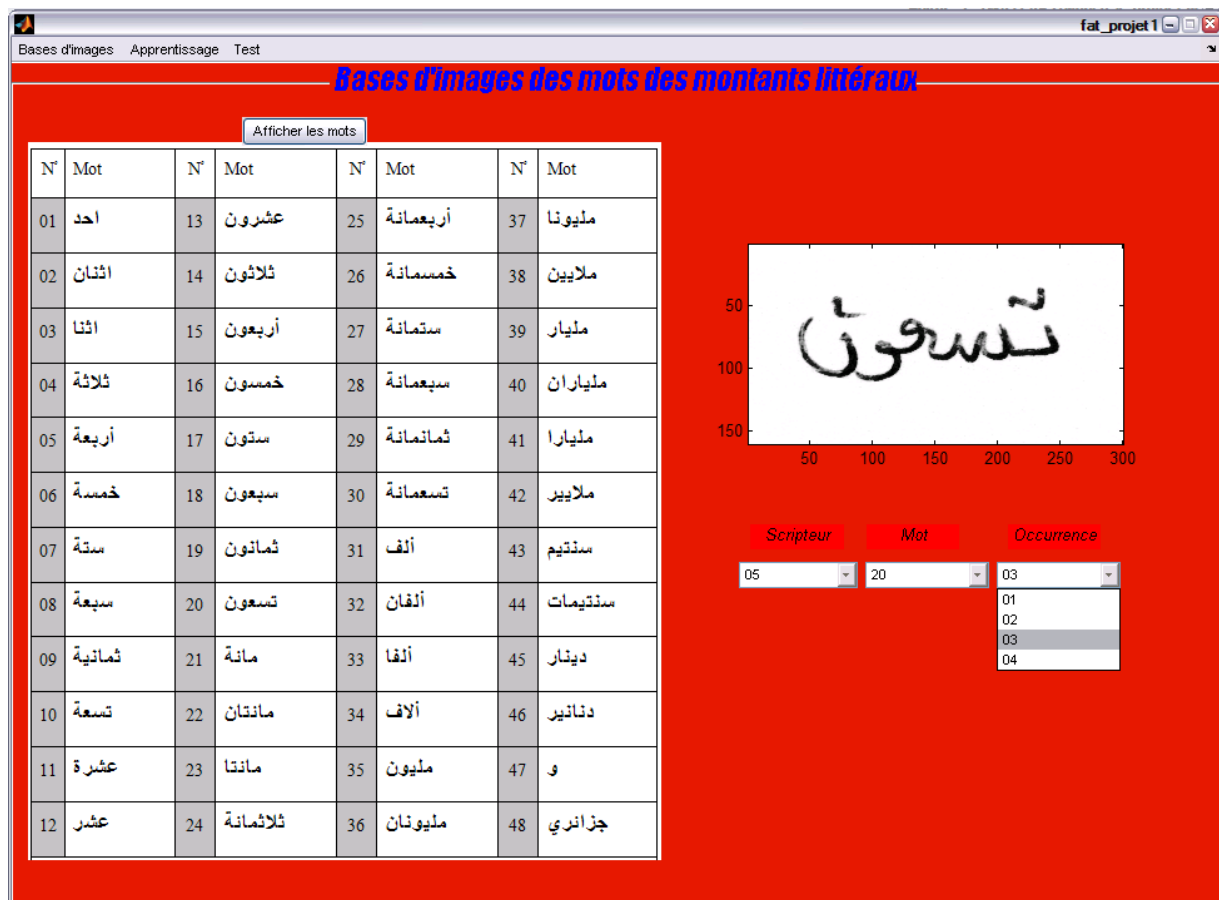


Figure2 : La base d'images des mots des montants littéraux arabes.

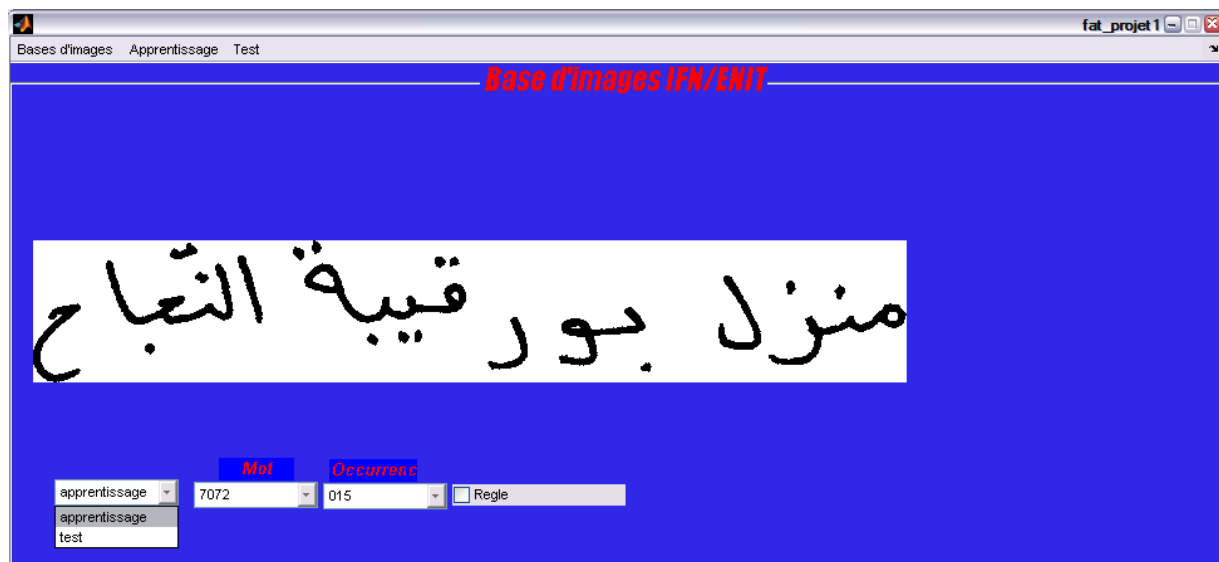


Figure3 : La base d'images IFN/ENIT.

- Dans le menu « Apprentissage », vous pouvez sélectionner la base d'images sur laquelle l'apprentissage sera effectué (voir figure4).



Figure4 : Le menu « Apprentissage ».

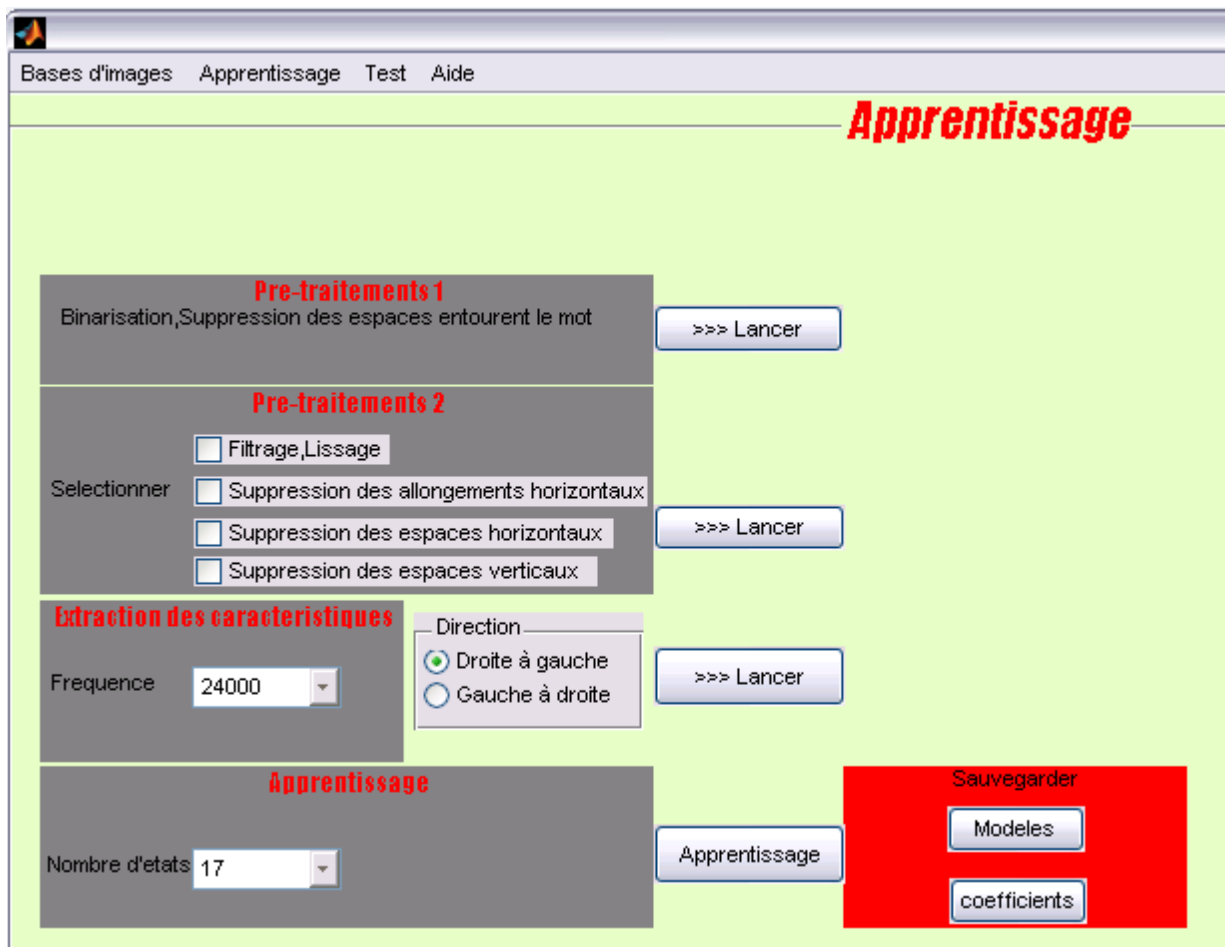


Figure5 : Fenêtre d'apprentissage de la base d'images des mots des montants littéraux arabes.

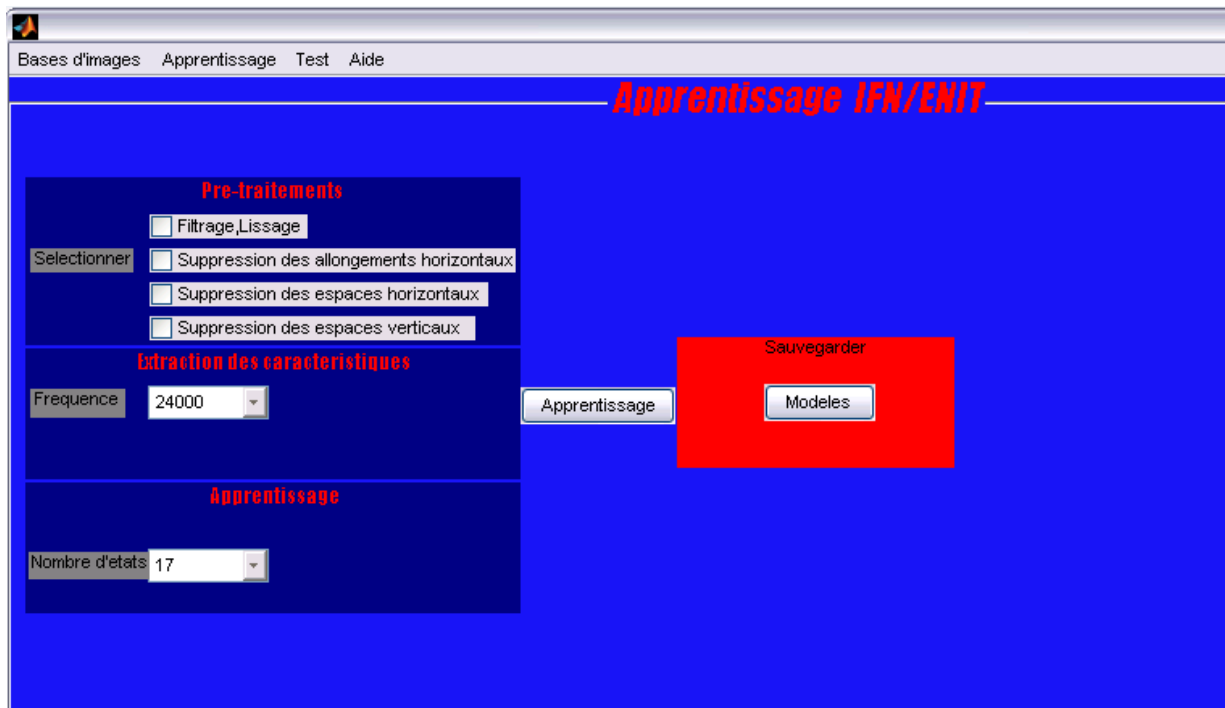


Figure6 : Fenêtre d'apprentissage de la base d'images IFN/ENIT.

- Dans le menu « Test », vous pouvez sélectionner la base d'images sur laquelle le test sera effectué (voir figure7).



Figure7 : Le menu « Test ».

Après la sélection de la base d'image, une fenêtre de test est affichée. Dans cette fenêtre vous pouvez choisir un mot à tester en cliquant sur le bouton « Ouvrir », l'image du mot choisi sera ensuite subie à une série d'opérations de prétraitements (Binarisation, PRE1=Filtrage,

PRE2=lissage, Suppr_allonge_H, Suppr_V_H, Suppr_V_V et Extraire) avant d'être reconnue en cliquant sur le bouton « Reconnaître » (voir figure8).

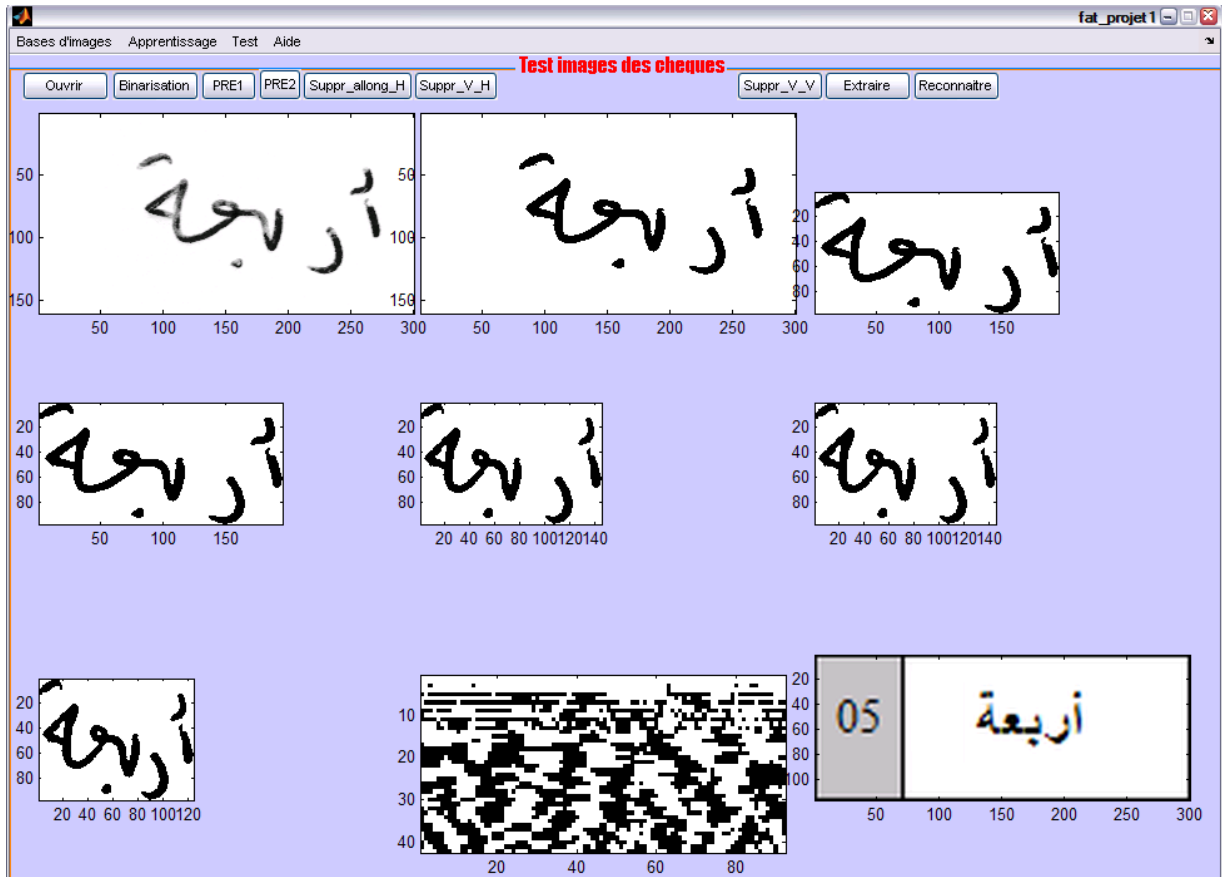


Figure8 : Fenêtre de Test du mot « أربعة » de la base des montants littéraux.



Figure9 : Fenêtre de Test du mot « زنوش » de la base IFN/ENIT.

- Dans le menu « Aide », vous pouvez consulter le manuel d'utilisation du logiciel (voir figure10).

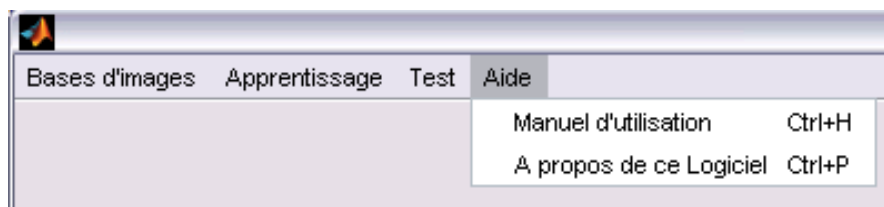


Figure10 : Le menu « Aide ».