Off-line Arabic Handwritten OCR Based on Web Services

By

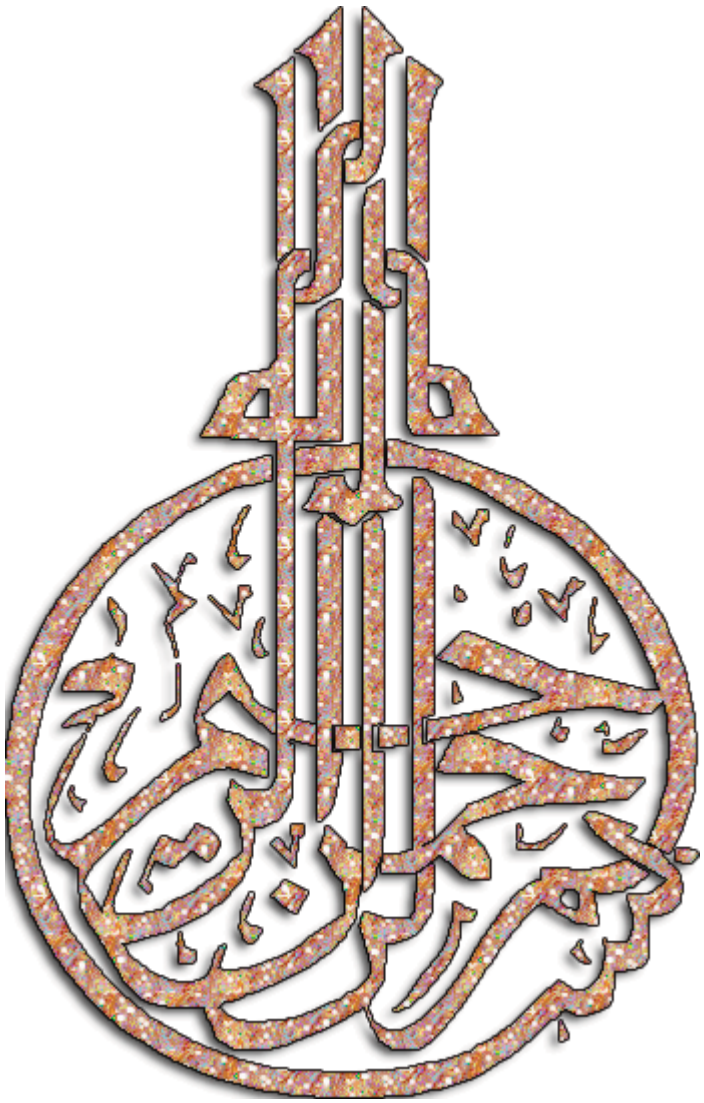
Khalil Ahmad Yassen Al-Sulbi

FACULITY OF COMPUTING AND INFORMATION TECHNOLOGY

KING ABDUL AZIZ UNIVERSITY

JEDDAH – SAUDI ARABIA

Shaaban 1436 H – 25 May 2015 G



Off-line Arabic Handwritten OCR Based on Web Services

By

Khalil Ahmad Yassen Al-Sulbi

A thesis submitted for requirements of the degree of Master of Science

(computer science)

Supervised By
Dr.Maher Khmakhem

FACULITY OF COMPUTING AND INFORMATION TECHNOLOGY

KING ABDUL AZIZ UNIVERSITY

JEDDAH – SAUDI ARABIA

Shaaban 1436 H – 25 May 2015 G

Off-line Arabic Handwritten OCR Based on Web Services

By

Khalil Ahmad Yassen Al-Sulbi

This thesis has been approved and accepted in partial fulfillment of the requirements for the degree of Master of Science (computer science)

EXAMINATION COMMITTEE

	Name	Rank	Field	Signature
Internal Examiner	Dr.Anas Fattoh	Associate Professor	Automatic Control	
External Examiner	Dr.Hani brdesee	Assistant Professor	electronic commerce	
Advisor	Dr.Maher Khmakhem	Associate Professor	Pattern recognition	

KING ABDULAZIZ UNIVERSITY
Shaaban 1436 H –25 May 2015 G

Dedicated to

My parents, my brothers, my sisters and my friends

ACKNOLEDGEMENT

At the beginning praise and gratitude be to ALLAH almighty, without His gracious help, it would have been impossible to accomplish this work. Working on this thesis has been a very interesting and valuable experience to me and I have learned a lot. I want to express my thanks to the people who have been very helpful during the time it took me to finish this thesis.

First, I would like to thank my supervisor professor *Dr.Maher Khemakhem* who helped me with guidance, supervision and constructive comments until I completed this work. He is not only my supervisor; he is just like my brother. I enjoyed working on this thesis under his supervision, I am grateful to him.

My special gratitude goes to my parents whose love and affection is the source of motivation and encouragement for my studies. I would like to thank all my family for unconditional support, and being there for me at any time.

I would like to express my gratitude to all my colleagues and my friends for useful discussions and prayers, in addition to everyone who asked, supported and encouraged me in any way.

Thank you all…
Khalil Ahmad Al-Sulbi, May, 2015

ABSTRACT

Nowadays, a large quantity of handwritten Arabic documents are waiting for their computerization since they are very rich and useful for several purposes. The manual computerization of such documents is very expensive in time and cost.

Several researcher in Arabic optical character recognition attempted to developed approaches and techniques for the automatic computerization of such documents, unfortunately and despite the diversity of these approaches and techniques the problem remind unsolved yet. Because of the difficulty of the morphological structure of the Arabic written.

In this thesis, we study and compare in depth some promising approaches and techniques in order to select some complementarity techniques or approaches to exploit the complementarity of these approaches, we present a hybrid K-NN/SVM for Arabic character recognition and build them as web services for improving the recognition rate through their cooperation. In addition, our approach offers a number of benefits: in the first hand, the hybrid KNN-SVM classifiers showed an improvement in terms of recognition rate. In the second hand, building these system in the form of web services helped us to benefit from the cooperation methods that have been selected, also makes our system available for use by other applications on the web.

المستخلص

يوجد بالعديد من المكتبات العربية كم هائل من الوثائق المكتوبة بالخط اليدوي والتي تمثل ثروة قومية من الواجب استغلالها لما تحتويه من معلومات مهمة . وتعتبر حوسبة هذه الوثائق من الخطوات الهامة للاستفادة منها ، إذ يصبح بالإمكان معالجتها بشتى الوسائل المتاحة لاستخراج ما فيها من معلومات مفيدة .

الحوسبة اليدوية لهذه الوثائق مكلفة جدا وقتا ومادة ، لذلك انكب بعض الباحثين لدراسة إمكانية الحوسبة الأوتوماتيكية لهذه الكتابة شأنها شأن الكتابات الأخرى خاصة اللاتينية. حيث أظهرت النتائج ضعفا واضحا لكل هذه التقنيات رغم تعددها واختلاف تعقيدها.

في هذه الأطروحة، تمت دراسة بعض الطرق المستخدمة في عملية التعرف على الحروف العربية ومقارنتها، لمحاولة تطوير أدائها باستخدام تعاون بعض هذه التقنيات المكلمة لبعضها وبناءها على شكل خدمات ويب لتيسير وظائفها واستعمالها.

تم اختيار الطريقتين (K-NN) و (SVM) وبناءهما على شكل خدمات ويب للاستفادة من تعاون هاتين الطريقتين في تحسين معدل التعرف للحروف العربية، وأظهر النظام الهجين (K-NN/SVM) تحسنا في معدل التعرف وكذلك ساعدتنا خدمات ويب في جعل التطبيق متاح للاستخدام من خلال تطبيقات اخرى خاصة على شبكة الإنترنت.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1 : Introduction

Chapter 2 : Literature review

LIST OF TABLES

Chapter 2 : Literature review

Chapter 3 : Off-line Arabic Handwritten OCR Based on Web Service (OAHOCRWS)

Chapter4 : Off-line Arabic Handwritten OCR Based on Web Service (OAHOCRWS)

LIST OF ABBREVIATION

OCR	Optical Character Recognition
XML	Extensible Markup Language
OAHOCR	Off-line Arabic Handwritten Optical Character Recognition
OAHOCRWS	Off-line Arabic Handwritten OCR based on Web Services
ICFHR	International Conference of Frontiers in Handwriting Recognition
ICDAR	International Conference on Document Analysis and Recognition
UDDI	Universal Description, Discovery, and Integration
WSDL	Web Services Description Language
URL	Uniform Resource Locator
HTTP	Hypertext Transfer Protocol
SOAP	Simple Object Access protocol

BPEL	Business Process Execution Language
WSCDL	Web Services Choreography Description Language
WSMO	Web service Modeling Ontology
OWL-S	Web Ontology Language
QoS	Quality of Services
HMM	Hidden Markov Model
K-NN	K-Nearest Neighbors algorithm
NN	Neural Network
SVM	Support Vector Machines
MLP	Multi-Layer Perceptron
RBF	Radial Basic
IFN/ENIT	Database of handwritten Arabic word
WS	Web Service
DSOA	Demands aware Service-Oriented Architecture
SOC	Service Oriented Computing
P2P	Peer-to-Peer
WSC	Web service composition

SOA	Service-oriented architecture
D&C	Divide-and-Conquer
PAW	Piece of Arabic word

Chapter 1

Introduction

1.1 Overview

This chapter provides an introduction to this research. It provides an overview about the problem statement and a general background about Arabic optical recognition objectives and motivation of this research and research methodology. A brief description of the research organization is provided towards the end of this chapter.

1.2 Problem statement

Human writing can be hard to understand or to recognize by a mechanical and automated system. In order to simplify the process, it is needed for the specific software to work using a worldwide web service database, where every character is associated to a typed character, no matter the handwriting style of the user. Off-line recognition of handwriting can be a very difficult task, since people have different styles of writing using their hands rather than typing the text using a physical keyboard.

The objective of this thesis is to improve the recognition rate of the Off-line Handwritten Arabic Optical Character Recognition .

1.3 Background

Optical Character Recognition

Optical character recognition (OCR) is a computer software designed to translate images of typewritten or handwritten text into machine-editable text encoded in a standard encoding scheme [1].

In this work we will attempt to improve the recognition rate of the Off-line handwritten through the adaptation, the improvement and maybe the cooperation of some strong complementary approaches and techniques. such a cooperation can be achieved easily through web services which present several advantages especially the possibility of combining (composing) such approaches and techniques.

Web Services

A web service is a software that is available over the internet and uses a standardized XML messaging system. All communications to a web service are encoded by XML, for example, a client request a web service by sending an XML message, then waits for a corresponding XML response[2].

Web Services Composition

Provides an open, standards-based approach for connecting web services together to create higher-level business processes [3].

1.4 Thesis Motivation

Arabic text recognition is very important subject for both Arabs and non-Arabs. Arabs need these tools for transforming all printed and handwritten materials into electronic media, while non-Arabs look at it as intermediate step towards fully transliterating Arabic text particularly old Arabic manuscripts.

There is a very rich collection of Arabic manuscripts around the world. Scholars need such text in machine readable from Arabic sources.

1.5 Thesis Objectives

A long-term goal is automatic transliteration of Arabic handwritten script. The aim of this thesis is to improve the performance of some approaches of Arabic handwritten OCR by implementing them as web services which can cooperate together. There are also some underlying goals which harmonies with the main objective :

- Studying the existing approaches related to offline Arabic handwriting OCR.
- Comparing and selecting some strong complementary approaches.
- Trying to adapt and improve them after that implementing them as web services which can cooperate together.
- Evaluating and adjusting the generated web services.
- Writing up the report.

1.6 Thesis Methodology

The methodology will be as follows:

- First, we have started in planning and collecting enough knowledge about the area of this research and have understood how Arabic handwritten OCR methods work. Then we have studied the advantages and the disadvantages of these methods, and tried to understand the problems which are not yet solved. Finally we have selected some strong approaches and especially complementary ones.
- Second, is the design stage which is the most critical task, we have designed the new OAHOCR (off-line Arabic Handwritten Optical Character Recognition) approach, that is based on the cooperation of the complementary approaches.
- Third, the implementation part: we have implemented all the tasks that have been clarified in the design stage.
- Fourth, results comparison and evaluation process have been done and reported.
- Finally, we have written the thesis report.

1.7 Thesis Organization

Thesis structure is organized as follows :

Chapter One : *"Introduction"* presents a brief overview of this study and includes the research background, research problem, research objectives, research methodology, and a basic organization of the thesis.

Chapter Two : "*literature review"* presents related works of this research and literature review of Arabic handwritten OCR, web service and web services composition.

Chapter Three : "*OFF-LINE ARABIC HANDWRITTEN OCR BASED ON WEB SERVICES (OAHOCRWS)"* presents the architecture and design of off-line Arabic handwritten OCR based on web services (OAHOCRWS**).**

Chapter Four : "*IMPLEMENTATION AND PERFORMANCE EVALUTION OF (OAHOCRWS)"* discusses the implementation details and performance evaluation of (OAHOCRWS).

Chapter Five : "*CONCLUSION AND FUTURE WORK"* concludes this research ideas and points out the limitations, advantages and the future works that can be done to extend the scope of this research.

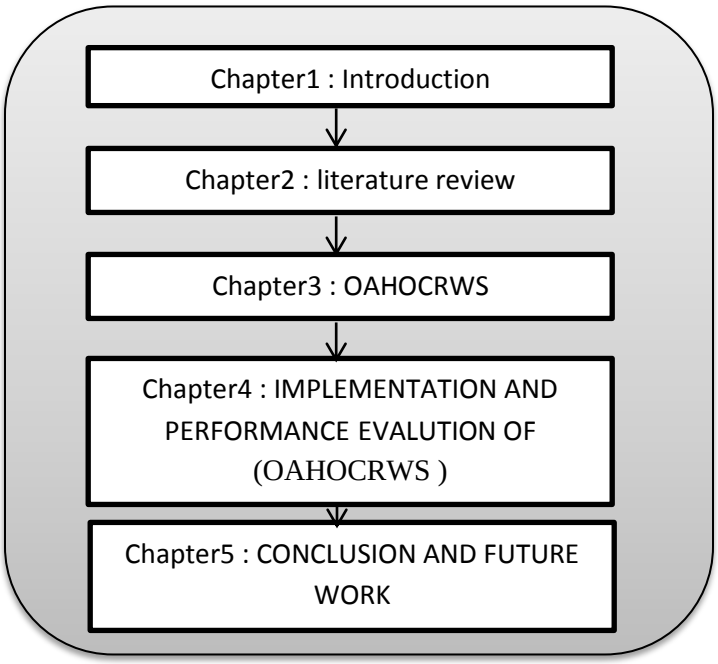


Figure 1.1 : Thesis Organization

Chapter 2

Literature Review

2.1 Overview

In this chapter we try to give a state of the art related to Arabic handwritten OCR, web service and web services composition.

2.2 Handwritten Optical Character Recognition

2.2.1 Definition of Handwritten OCR

There are many specialized conferences in this field of research such as, the International Conference of Frontiers in Handwriting Recognition(ICFHR) and the International Conference on Document Analysis and Recognition (ICDAR).

Handwriting recognition software insists on making sure that all written characters will have a corresponding from the language database that is being used by the system. Among the easiest languages to recognize comes the English language, compared to Arabic handwriting. The International Arab Journal of Information Technology underlines the fact that “in contrast to Latin script, the recognition of Arabic characters is relatively more difficult due to its inherent characteristics. The crucial step in recognizing Arabic characters is the

segmentation of word into separate characters. The variable size of Arabic characters is one of the reasons that make the segmentation and recognition hard. Baseline detection is also a problem that makes the segmentation step difficult in machine printed as well as handwritten Arabic text [4].

As the Universal Journal of Computer Science and Engineering Technology underlines it, “Optical Character Recognition (OCR) is the mechanical or electronic translation of scanned images of handwritten, typewritten or printed text into machine-encoded text”. It is widely used to convert books and documents into electronic files, to computerize a record-keeping system in an office, or to publish the text on a website.

2.2.2 Types of OCR

Character recognition software uses two different methods to input the handwriting text that will be recognized: on-line and off-line as illustrated in Figure (2.1).

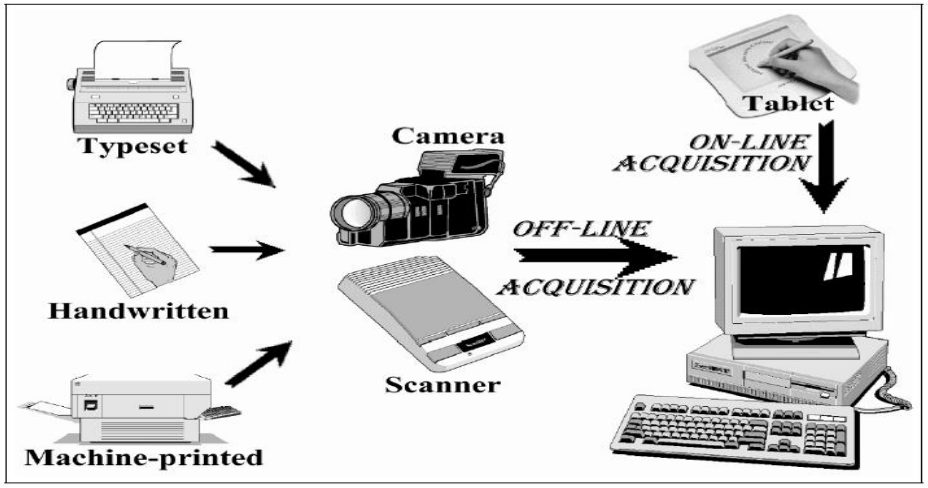


Figure 2.1 : Difference between off-line and on-line[5]

On one hand, the on-line systems recognize the text by the use of temporal dynamics of the information written, recognizing either one single character at a time or the entire word once the text is fully written. The information that is input in the system contains the duration, number and order of each stroke, mostly using a touchscreen device like a tablet or a smartphone. This specific type of input is also very common at banks or local businesses, in order to avoid having the signature on paper, which can be copied easily by anyone who reaches this information[6,7].

On the other hand, off-line softwares work with scanned images or photographs, from pages that will be converted to a binary image. These systems do not have access to the same information like the on-line softwares do, therefore off-line handwriting recognition is sometimes more challenging than on-line systems[8].

2.2.3 General stages of character recognition system

The process of optical character recognition of any script can be broadly broken down into five stages: Pre-processing, segmentation, feature extraction, classification and post-processing[9] as illustrated in Figure(2.2).

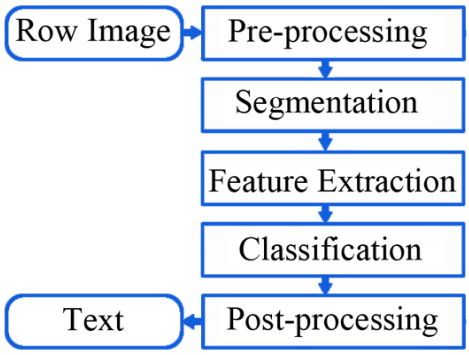


Figure 2.2 : Stages of character recognition system[9].

- **Pre-processing :**

  Pre-processing, to begin with, endeavors to produce data that user friendly to the OCR systems thereby making it easier for the system to accurately and smoothly operate. Pre-processing operates on the following k mandates to ensure efficiency; noise reduction, skew correction, binarization and strike width normalization.

- **Segmentation :**

  The goal of segmentation step is to partition word image into components. Each component containing an isolated complete character. In Arabic handwritten word is extremely difficult to segment for several reasons, the most important of these reasons is the connected nature of Arabic words. Another reason is that sometimes one letter appears above or below the previous letter [1]. Sometimes the document contains both text and graphics, in this case segmentation occurs on two levels. On the first level, text and graphics are first separated. Then the text is separated into its components down to individual characters.

- **Feature Extraction :**

  Feature extraction aims at representing each character by an invariant feature vector, an entity that eases and aims at maximizing recognition rate with least data quantity. This stage is based on three subcategories which include structural, statistical and global conversion and moments.

- **Classification :**

The classification is the main decision making stage in OCR system. The feature of the image are extracted and given as input to the classifier. classifier compare the input feature with stored pattern and find out the matching class for input.

In general there are two basic strategies for recognizing words: Holistic strategies or Analytic strategies.

1-Holistic strategy recognizes the whole words or sub words without requiring segmentation, but most of proposals work on limited vocabularies.

2-Analytic Strategy recognizes the segmented features, so it requires segmentation, and can be applied on unlimited vocabularies.

Table (2.1) classifies text recognition methods according to segmentation use or not.

Table 2.1 : Holistic and Analytic Approaches

Holistic strategy	**Analytic Strategy**
Amrouch et al.(2011),(decision tree) [10]	Jafaar alabodi and Xue Li, (2013)[11]
Jawad et al.(2011),(HMM)[12]	Alaei et al.(2010),(SVM) [13]
N. AZIZI et al. (2010),(HMM) [14]	Kumar et al.(2010),(SVM)[15]
Farah 2004,(ANN, KNN & Fuzzy KNN) [16]	Laslo Dinges et al.(2011),(K-NN) [17]

Farah 2005,(MLP)[18]	Farah,(2012),(Multi-NN)[19]
Ahmed and Suen,(2014),(novel)[20]	Moncef Charfi et al.(2012)[21]
Pechwitz 2003,(HMM)[22]	Zaiz Faouzi et al.(2010),(SVM) [23]
Abdelwadood Mesleh et al.(2012)[24]	Haidar Almohri et al.(2008),(Fuzzy ART Neural Network)[25]

- **Post-processing :**

  Lastly, the post-processing phase which is apparently the last phase in the series, it operates on such mandates as to improve recognition rate through refinery of decisions.

2.2.4 Difficulties from Characteristics of the Arabic Writing System

As mentioned in[26] , the main characteristics of Arabic writing can be summarized as follows:

- Arabic text (machine printed or handwritten) is written cursively and in general from right to left. Arabic letters are normally connected to the baseline.
- Arabic writing uses letters (which consist of 28 basic letters), ten Hindi numerals, punctuation marks, spaces, and special symbols.
- Arabic letters have up to four different shapes, depending on its position in the word, this increases the number of classes from 28 to 100 (Table 2.2). For example the letter (ع) has four different shapes, at the

beginning of the word, the middle, and the end of the word and one in isolation as a standalone word as illustrated in Figure(2.3).

Figure 2.3 : Different shapes of the Arabic letter (ع) in:
(a) beginning, (b) middle, (c) final and (d) isolated[26].

Table (2.2) : Arabic alphabet in all its forms

Name	Isolated	Start	Middle	End
Alif	ا			ـا
Ba	ب	بـ	بـ	ـب
Ta	ت	تـ	تـ	ـت
Tha	ث	ثـ	ثـ	ـث
Jeem	ج	جـ	جـ	ـج
Hha	ح	حـ	حـ	ـح
Kha	خ	خـ	خـ	ـخ
Dal	د			ـد
Thal	ذ			ـذ
Ra	ر			ـر
Zay	ز			ـز
Seen	س	سـ	سـ	ـس
Sheen	ش	شـ	شـ	ـش
Sad	ص	صـ	صـ	ـص
Dhad	ض	ضـ	ضـ	ـض
Tta	ط	طـ	طـ	ـط
Za	ظ	ظـ	ظـ	ـظ
Ain	ع	عـ	عـ	ـع
Gain	غ	غـ	غـ	ـغ
Fa	ف	فـ	فـ	ـف
Qaf	ق	قـ	قـ	ـق
Kaf	ك	كـ	كـ	ـك
Lam	ل	لـ	لـ	ـل
Meem	م	مـ	مـ	ـم
Noon	ن	نـ	نـ	ـن
Ha	ه	هـ	هـ	ـه
Waow	و			ـو
Ya	ي	يـ	يـ	ـي

- Arabic uses diacritical markings (Table 2.3), The presence or absence of diacritics indicates different meanings of the same word. Table (2.4) gives an example of an Arabic word with different diacritics indicating four different meanings.

Table (2.3) : Diacritical markings in Arabic written[26] .

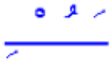
Diacritics	**Figure**
Single diacritics	
Double diacritics	
Shadda	
Combined diacritics	

Table (2.4) : Example of an Arabic word with different diacritic indicates different meanings[26].

Arabic word	**English meaning**
دَرَسَ	he studied
دَرَّسَ	a lesson
دَرَّسَ	he taught
دُرِسَ	it was studied

- Arabic writing contains many fonts and writing styles, also characters of the same font have different sizes as illustrated in Figure (2.4), Hence, segmentation which is based on fixed size or width cannot be applied to the Arabic language.

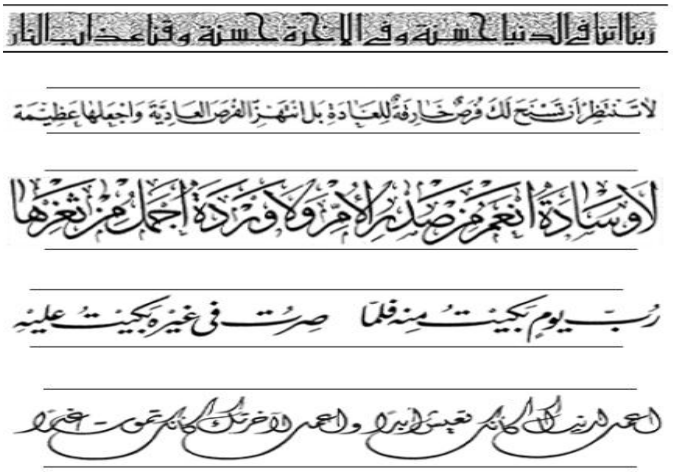


Figure 2.4 : Different Arabic sentences in different styles[26].

- Arabic uses Ligatures, Ligatures are combinations of two, or sometimes three characters into one shape (Figure2.5). Ligature selection is dependent not only on the characters themselves but also on the selected Arabic font.

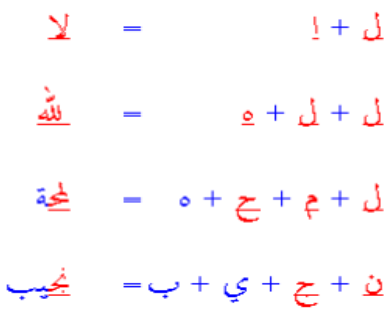


Figure 2.5 : Arabic Ligatures[26] .

2.3 Previous studies of Arabic OCR :

Arabic handwriting OCR (AHOCR) is very important since a large amount of Arabic documents are still waiting for their computerization due to the richness of their contents. The manual computerization of these documents will be very expensive in cost and time. A lot of research has been done in this area, but researchers still continue to optimize the recognition rate which is yet low in AHOCR compared with other languages such as Latin. We will review and summarize next some of existing works in AHOCR.

Arabic handwritten OCR:

The following table shows a summary of some of the previous studies in the field of Arabic handwritten OCR.

Table 2.5: some of the previous studies in the field of Arabic handwritten OCR

Author	Feature	Classifier	Size of word	Rec. Rate (%)
Amrouch et al(2011) [10]	Structural	HMM	48 city names written three times	75.74
Safabakhsh (2005) [11]	Fourier descriptors, number of loops, high/width ratios, pixel densities, positions of right and left connection	continuous – density variable-duration HMM	50 words	90.48
Alaei et al. (2010) [13]	modified chain code directional frequencies	SVM	15338 Characters	96.68
Kumar *et al.* (2011) [15]	intersection and open end points features	SVM	not mentioned	90
Farah et al (2004)	Structural	ANN, KNN & Fuzzy KNN	48 words (100 writers)	96

[16]				
Pechwitz and margner (2003) [22]	Columns of pixels in the blurred thinned image	HMM	26,459 images of 946 city names	89
Dutta and Chaudhury (1993) [27]	curvature features	feed-forward NN	12000 words	85
G S Lehal et.al (2000) [28]	Contour Extraction	MQDF	Number of words 8,43,590 Number of Characters 32,72,268 Number of Unique words 55,071	89
Bennamoun and Bergmann (2000)[29]	chain code	Recognition based-segmentation	20 characters	90
Dehghan et al. (2001) [30]	Histograms of freeman chain code	HMM	17,000 images of 198 names	65
Khorsheed (2003) [31]	Structural	HMM	405 characters	87
Alma'adeed et al.(2004) [32]	Structural	Set of HMMs	4700 words	60
Souci et al (2004) [33]	Structural	NN	55 words	92
El-Hajj (2005) [34]	Densities of foreground pixels, concavity and derivative features in a sliding window	HMM	21500 words	87.7
Farah et al. (2005) [35]	Structural and Statistical	ANN Multi classifiers	2400 words	95.2
Mazaffari et al (2005) [36]	Average variance of X and Y change in portions of the skeleton	nearest – neighbor	2880	87.26
Pechwitz et al (2003)	Skeleton directions, Pixel values	SC-1D-HMM	training: a-c, test: d	89.1

[37]				
Pal *et al.* (2006) [38]	Gradient features	Unknown	training: a-e, test: f	83.34
Menasri et al.(2007) [39]	Graphemes	HMM/NN	training: a-e, test: f	87
Touj et al. (2005) [40]	Directional values, connection of graphemes	Planar HMMs	training: a-d, test: e	86.1
Hani Khasawneh (2006)[41]	silent features	Neural Network	5464	87
Rajashekararadhya *et al*. (2008) [42]	zoning features	SVM	Unknown	98.6
Dreuw et al. (2008) [43]	Image slices and their spatial derivatives	HMM	training: a-d, test: e	80.95
Natarajan et al.(2008) [44]	Percentile of intensity, energy, correlation, angle	HMM	Unknown	89.4
Benouareth (2008) [45]	Distribution, concavity and skeleton features	HMM	training: a-c, test: d	89.08
Patil and Ramakrishnan (2008) [46]	Gabor and Discrete cosine transform (DCT)	KNN and SVM	20000 words	89
Moussa et al (2008) [47]	Fractal analysis features	KNN Classifier and Radial Basic Function (RBF)	Arabic and Latin	96.64 % (KNN) and 98.72 % (RBF)
Desai (2010) [47]	features abstracted from four different profiles of digits	neural network	2650 digits	82
Sarkar et al (2003) [48]	Horizontalness feature, segmentation-based feature and foreground-background transition feature	Multilayer perceptron. Trained with back propagation algorithm	3050 characters	98.43
Jafaar Alabodi	junction-points, end	Unknown	2000 words	93.3

and Xue Li (2013) [51]	points, and connectivity between the components for the further processing needs			
Abdurazzag and Salem(2007) [52]	Structural	Wavelet Compression	1968 letters	80
Almuallim and Yamaguchi (1987) [53]	geometric and topological	String matching	400 words	91

In [16], Farah et al presented an new Arabic OCR system that used three classifiers (ANN, KNN and Fuzzy KNN) and structural features. The classification step performed in parallel combination schema. The result of classification stage normalized first, and combination schema is performed. At last the decision on Specified words can be done. This system tested in 48 words achieved a 96% recognition rate.

In [29], Bennamoun and Bergmann proposed an Arabic OCR system that uses recognition-based segmentation technique to overcome the problem of the classic segmentation. This system is composed of five stages and feedback loop (Figure 2.6 [8]), the five stages are (1) image acquisition, (2)preprocessing, (3) word segmentation and character fragmentation, (4) feature extraction and (5) classification. The feedback loop, which is linked between the segmentation and recognition stages, conducts a signal to control the combination of character fragments. The system was implemented and the results show a 90% recognition rate.

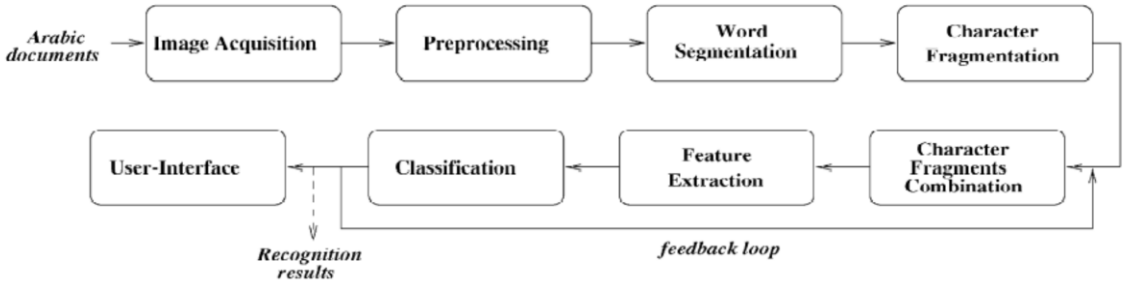


Figure 2.6 : The structure of the proposed Arabic OCR system

In [30], Dehghan et al presented an new method for off-line recognition of isolated handwritten character based on HMM classifier and used methods based on regional projection contour transformation (RPCT) for features. This system consists of two phases: training and recognition phases as illustrated in figure (2.7). This system tested on 198 names of cities and achieved a 65% recognition rate.

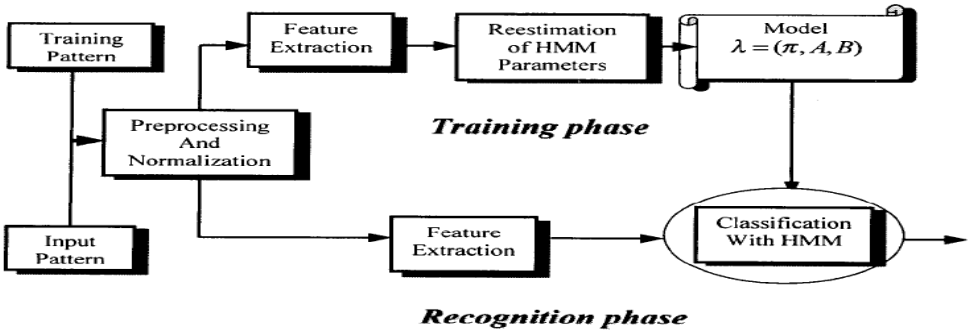


Figure 2.7 : Overall structure of the proposed handwritten character recognition system

In [31], Khorsheed presented an new Arabic OCR system based on HMM and structural features. The structural features for the handwritten script were

extracted after decomposing the word skeleton into a sequence of links with an order similar to the writing are. Using the line approximation, each line was broken into small line segments, which were transferred into a sequence of discrete symbols using vector quantization(VQ). Then an HMM recognizer was applied with image skeletonization to the recognition of an old Arabic manuscript (Khorsheed, 2000). The HMM was performed using 296 states on 32 character models, and each model was left to right HMM with no restriction jump margin. The system achieved a 89% recognition rate.

In [37], Pechwitz and margner also presented an new Arabic OCR system based on HMM classifier and used columns of pixels in the blurred thinned image as features. the system achieved a 89% recognition rate using IFN/ENIT database (26,459 images of Tunisian city-names).

In [39], Menasri et al proposed a new off-line handwritten Arabic words recognition system that used hybrid recognizer HMM/NN. In this hybrid system HMM model represent each letter-body class and NN computes the observations probability distribution. The authors tested their system using IFN/ENIT which has a vocabulary of 937 city names and achieved a 87% recognition rate.

In [40], the authors propose a planar modeling approach based on HMM. The aim of this approach is to decomposition the writing into a limited set of elementary entities. Therefor the writing was divided into five logical horizontal

bands corresponding to descanters, ascenders, median zone, upper diacritics and lower diacritics. Figure (2.8) shows how the segmentation process allows them to reduce the complexity of the treated shapes.

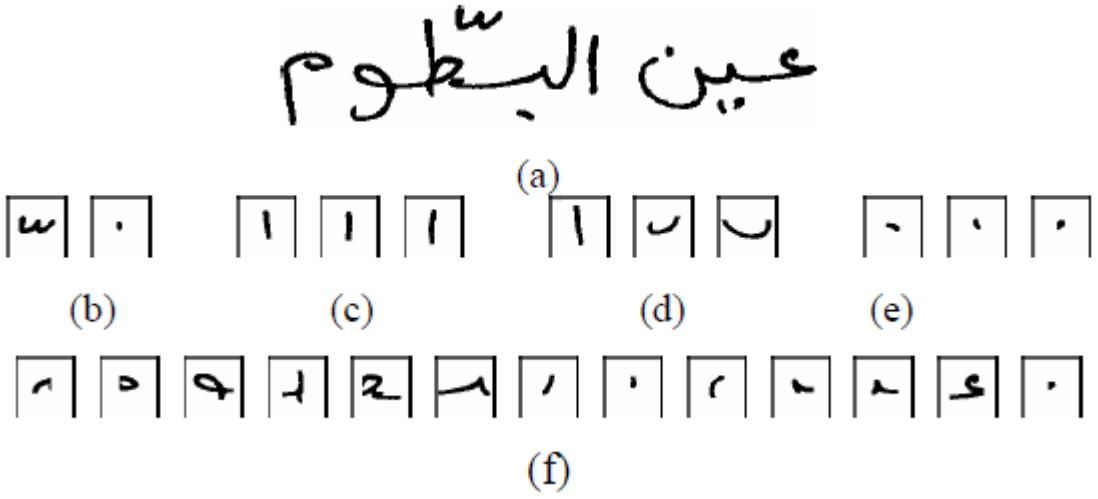


Figure 2.8 : (a) Original image of city name "عين البطوم" and the extracted grapheme belonging to, (b) upper diacritics, (c) ascenders, (d) descanters, (e) lower diacritics and (f) median zone.

The authors tested their system using IFN/ENIT and achieved a 86.1% recognition rate.

In [41], the author presented a new Arabic optical character recognition (*AOCR*) system Figure(2.9). The system accepts a scanned-page image containing a set of text lines, and affected by typical noise level. The system preprocesses the image before the separate lines are handled to the character extraction phase.

The neural network was used as classifier in this system and achieved a 87% recognition rate.

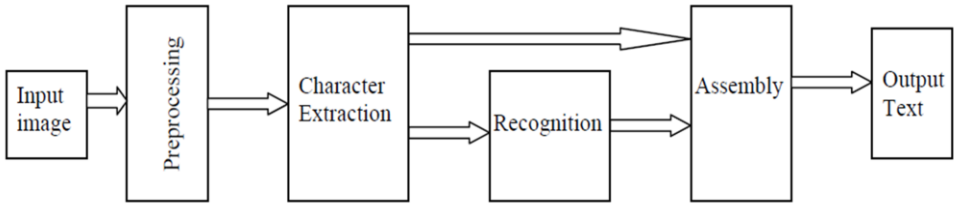


Figure 2.9: system architecture

In [43], the authors proposed an explicitly model white-spaces for Arabic handwriting recognition based on HMM. The proposed white-space character model uses a special and separate single-state HMM model with separate entry and exit penalties. The authors tested their system using IFN/ENIT and achieved a 80.95 % recognition rate.

In [50], the authors proposed an off-line Arabic handwriting recognition system. The processing is achieved in three main stages. Firstly, the image is skeletonized to one pixel thin. Secondly, transfer each diagonally connected foreground pixel to the closest horizontal or vertical line. Finally, these orthogonal lines are coded as vectors of unique integer numbers; each vector represents one letter of the word. the system has been tested on the IFN/ENIT database to evaluate the proposed techniques and achieved a 97% recognition rate.

The proposed algorithms are described according to the offline handwriting

recognition process described in Figure(2.10).

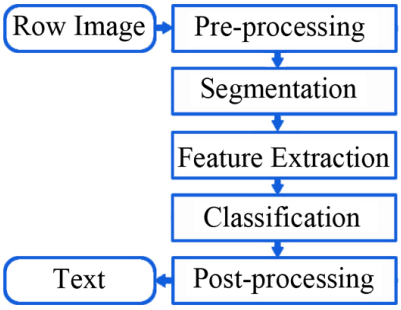


Figure 2.10: Stages of character recognition process.

In [52], Abdurazzag and Salem proposed an Arabic OCR system that consists of three stages, preprocessing, feature extraction and recognizer Figure(2.11). This research proposed a new construction technique of OCR similar to the one using wavelet compression. This system has achieved an accuracy (97.7% for some litters at average 80%).

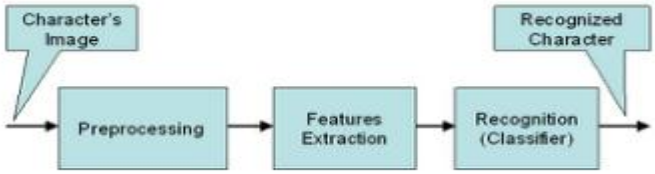


Figure 2.11: The typical structure of OCR system .

In [53], Almuallim and Yamaguchi proposed a system which consists of four stages, preprocessing, segmentation, feature extraction and classifier. In this research a structural recognition method of Arabic cursively handwritten words is proposed. At first the word is segmented into strokes, strokes with a loop and

strokes without a loop. These strokes are then further classified using their geometrical and topological properties. Finally, the relative positions of the classified strokes are examined, and the strokes are combined in several steps into the string of characters that represents the recognized word. This system was tested in 400 words and showed high recognition accuracy (91%).

In the next section we will give an overview on web services composition since we will use this technique to combine (integrate) OCR complementary approaches.

2.4 Web Services

To maintain a good management strategy in a given organization, they need various software systems in order to communicate. Web services offer methods of communication between two devices over a network, being described as "a software system designed to support interoperable machine-to-machine interaction over a network"[53]. Figure (2.12) shows web services architecture[54], this architecture sets forth three roles and three operations. The three roles are the service provider, the service requester, and the service registry. The objects acted upon are the service and the service description, and the operations performed by the actors on these objects are publish, find, and bind.

A service provider creates a web service and its service definition and then publishes the service with a service registry based on a standard called the Universal Description, Discovery, and Integration (UDDI) specification.

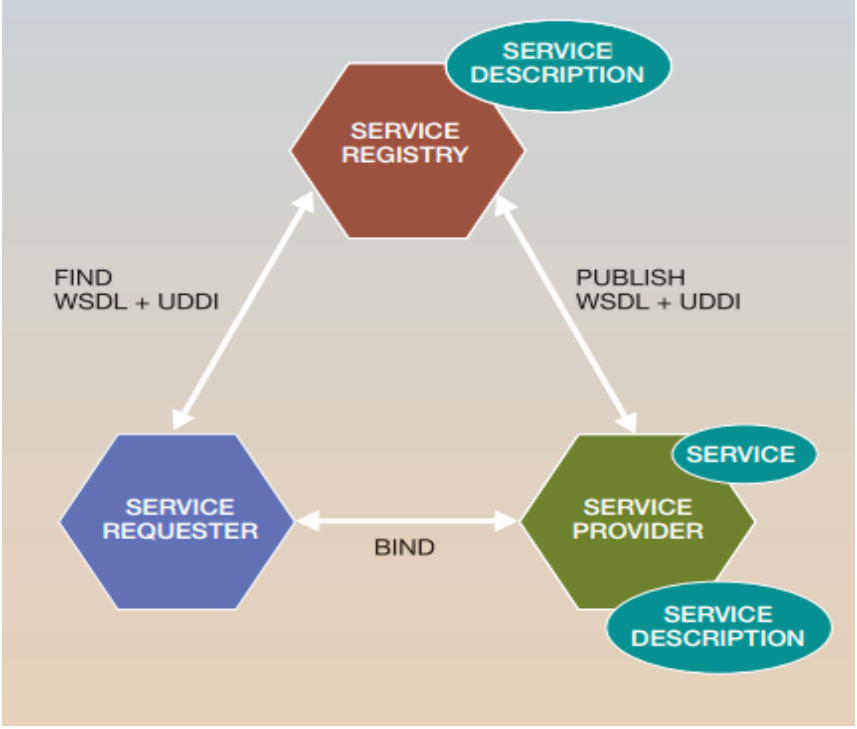


Figure 2.12 : web services architecture .

Once a web service is published, a service requester may find the service via the UDDI interface. The UDDI registry provides the service requester with a WSDL service description and a URL (uniform resource locator) pointing to the service itself. The service requester may then use this information to directly bind to the service and invoke it[53].

Web services are application components that communicate using open protocols, self-contained and self-describing. They can be discovered using UDDI and can be used by other applications, using HTTP and XML as basis[54]. Web services have two types of uses: reusable application components and connecting existing software.

Since they are operating through standards web environment like HTTP and XML-based protocols. Web services are powered by XML and work with WSDL, SOAP and UDDI. According to the World Wide Web Consortium[2], the eXtensible Markup Language (XML) is defined as a "meta-language for describing data" [55]. The data is being surrounded by text-based tags that are customizable and give information about the data and its structure.

The XML syntax consists of "text-based mark-up" that describes the data being tagged, it is both application-independent and human readable. This simplicity and interoperability have helped XML achieve widespread acceptance and adoption as the standard for exchanging information between heterogeneous systems in a wide variety of applications, including web services" (Cavanaugh, n.d.). The extensible markup language forms a base for modern web services in order to provide a description on their interface and also converting the messages into codes. The XML messaging are used for WSDL, SOAP and UDDI, so that any machine can interpret the language.

2.4.1 Web services benefits

For programmers and developers, the web services can offer several benefits in both technology and business. They use a versatile design, being able to be accessed by users through a client interface or through other web services or applications. Clients can use data from multiple services even if they are not compatible. The systems send information from one to the other through web service, so that if there's any change in the database it will not affect the service itself.

Another benefit is the application and data integration, so that the client can only need the WSDL definition to exchange data, no part needing to know the format or how it is implemented. “These benefits allow organizations to integrate disparate applications and data formats with relative ease”[56]. The exchange that is being done between the systems comes with a platform, an independent language, a vendor, XML technologies and HTTP as a transport, so that the web services can communicate better.

Using the code multiple times is another benefit that web services offer. It is a positive side-effect of their interoperability and flexibility, since the services can be used by many clients with different objectives. Therefore, there’s no need to create a customized service to each business, because portions of the service are re-used and simple.

Among the most important benefits we can notice that the use of web services saves money and cuts off most of the costs. It is easy to operate, having no customized data but codes that can be reused. The “investments in system development and infrastructure can be utilized easily and combined to add additional value. Since web services are based on open standards their cost is low and the associated learning curve is smaller than that of many proprietary solutions”[55]. Therefore, Web services have many advantages due to the protocols and the structure that they use for every organization, needing small investments.

2.4.2 Web services composition

Provides an open, standards-based approach for connecting web services together to create higher-level business processes[56].

Web service composition originated from the necessity to achieve a predetermined goal that cannot be realized by a standalone service. Internally, in a composition, services can interact with each other to exchange parameters, for example a service's result could be another service's input parameter. Figure (2.13) shows Design of a composite service .

The main purpose of web services is to integrate efficiently applications over the web[57].

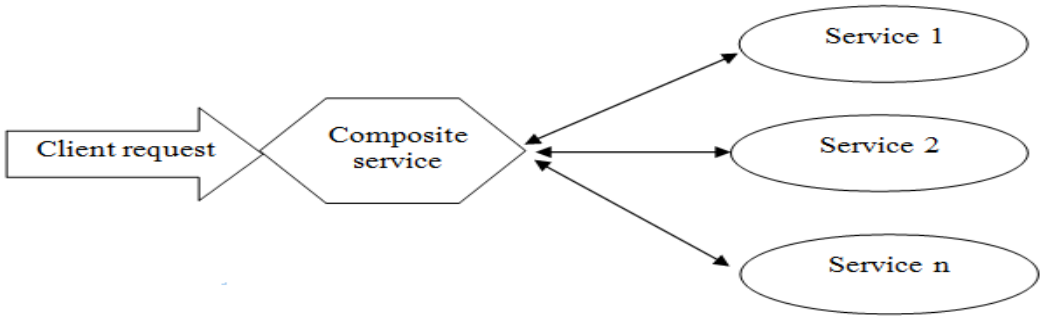


Figure 2.13 : Design of a composite service .

When it comes to the web services composition, we know that it’s studied to support business and enterprise applications to be integrated over web. The current web service composition “approaches range from practical languages aspiring to become standards (like BPEL, WSCDL, OWLS and WSMO) to theoretical models (like automata, Petri nets and process algebras)”[58].

Reusing the mechanism to build new applications is one of the main features in web services. There are multiple composition rules that need to be taken into consideration, like the syntactic and semantic web services composition.

1. *Syntactic web services composition*

   We have two approaches: an orchestration that combines web services through a central coordinator responsible for combining the activities, as well as the web services choreography that defines tasks which are complex, using the definition of the undertaken conversation by each participant. By peer-to-peer interaction among the web services collaboration, there are several proposals that exist to orchestrate the languages, one of them being BPEL.

2. *Semantic web services composition*

   Nowadays, technology is addressed towards the syntactic aspect of web services and provides a set of rigid web services that can't keep up with the environment that keeps changing. The semantic web services composition "provides a process-level description of web services which, in addition to functional information, models the pre- and post-conditions of processes so that the evolution of the domain can be logically inferred"[58]. The Internet is a big and complex database, where web pages have different structures, with powerful applications that are written in order to use annotations and engines to automatically execute scripts. In order for the web services composition to work, there

are two main initiatives to be taken into consideration: OWL-S and WSMO. OWL-S "defines the web services ontology with four main elements: the services, presents, described by and supports. WSMO introduces a set of core non-functional properties that are defined globally and that can be used by all its modeling elements"[58], one difference between them is the fact that OWL-S doesn't separate the user's need from what the service provides. At the same time, WSMO elements can have the non-functional properties expressed in any of them, while OWL-S uses vocabularies and restricts the properties to the service profile.

2.4.2.1 Web services composition characteristics

By studying the web services composition, we can observe some characteristics in report with the two composition rules. Therefore, we "believe that any web service composition approach should aim to support these characteristics, without pretending these to be all characteristics of importance"[58]. Among these characteristics we can talk about connectivity, exception handling, compensations, correctness and Qos.

Table 2.6 : Comparing standardization approaches[58].

	Syntax-based		Semantics-based	
Characteristics	**BPEL**	**WS-CDL**	**OWL-S**	**WSMO**
Connectivity	+	+	+	+
Exception handling	+	+	±	+
Compensations	+	+	-	-
Correctness	-	-	-	-
QoS	±	±	+	+

Connectivity needs to be reliable, to confirm the web services interactions before composition, so that the web service can have a continuous delivery afterword. Therefore, the service needs to be reliable, accessible; it should be able to handle exceptions and compensations. **Correctness** of the web service behavior is important when thinking that the composition of web services can lead to complex systems. The services need to be safe and correct, secure and trustful, as well as to satisfy the behavior needed. The third main characteristic is the **Quality of Services** (Qos), which shows how qualitative the web service is. It works on accuracy, availability and performance, to determine the success rate, measure response time, latency and overall time needed to process a request.

If we were to compare the standardization approaches, we see that for connectivity the approaches offer through the web service themselves, even if they are at a low

level and customized. However, correctness is not offered by any of the approaches through direct support to verify the web service composition. The quality of service requires “a careful consideration of the QoS characteristics of the constituent web services”[58].

2.4.2.2 Web services composition framework

Jinghai Rao and Xiaomeng Su[59], have proposed a general framework that can be adapted to web services. They consider that “it’s already beyond the human ability to analysis them [web services] and generate the composition plan manually” (n.d.). However, based on the general framework that they have developed we can observe that there are two types of participants: the service requester and the service provider. Between them, there are many types of processes and specifications, leading to the service requesters to consume the information received from the service provider Figure(2.14). The system also brings other components, such as the translator, process generator, execution engine, evaluator and service repository. The translator translates external languages used by the two participants, while the process generator generates a plan for the available services for the service repository to fulfill the requests. Through the entire process we can observe five main phases, related to each component from the system: presentation of single service, translation of the languages, and generation of composition process model, evaluation of composite service and execution of composite service.

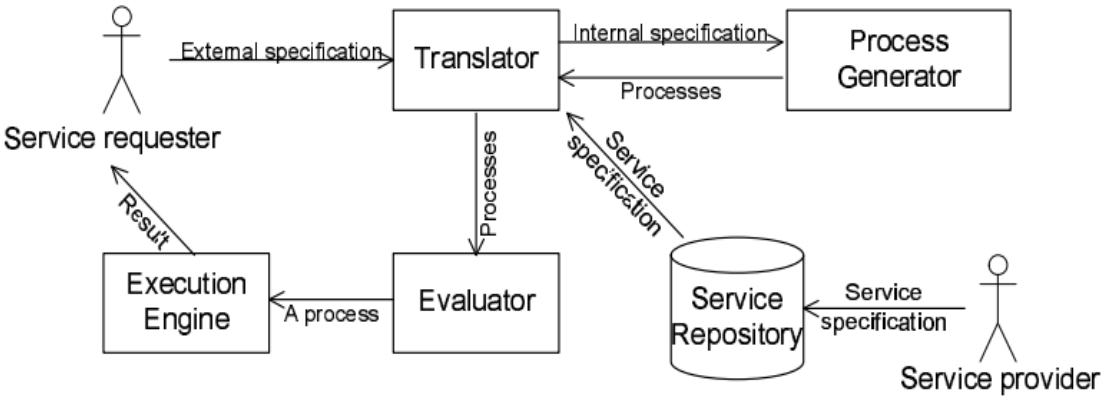


Figure 2.14 : Framework of the service composition system

Therefore, this web service composition framework consists of service presentation, translation, process generation, evaluation and execution. There are different platforms, methods and languages that are needed for each step. “The methods are enabled either by workflow research or AI planning. The workflow methods are mostly used in the situation where the request has already defined the process model, but automatic program is required to find the atomic services to fulfill the requirement. The AI planning methods is used when the requester has no process model but has a set of constraints and preferences, therefore the model can be generated by the program automatically. Even though automation can come in different methods and levels in service composition, the web service environment is complex and can’t generate everything automatically.

2.5 Previous studies of web services (composition) :

In [60], the authors shown that “The business world has developed a number of XML-based standards to formalize the specification of web services, their composition and their execution. On the other hand, the semantic web community focuses on reasoning about web resources by explicitly declaring their preconditions and effects with terms defined precisely in ontologies. Current service composition approaches range from practical languages aspiring to become industrial standards (e.g. BPEL and OWL-S) to more theoretical models and languages[61], in their article entitled A Survey on Service Composition Approaches: From Industrial Standards to Formal Methods. With the support of the “A. Faedo” Institution of Technology and Science of Information, they created a research that raises the need for “web service composition to provide the mechanism to fulfill the complexity of the execution of business processes”. Based on their research and analysis, they have created a survey to show the differences and similarities between formal methods and industrial standards. In their research they have shown that “an overview of service composition languages and models. Five approaches-namely BPEL, OWL-S, automata, Petri nets, and process algebras-were chosen and compared against a total of six requirements that a service composition approach should support in order to facilitate the composition of web services.

In [62], the authors have shown that “Web service composition (WSC) based SOA implementations have increasingly become a significant type of SOA

projects in practice. However, effort estimation for such a type of SOA project can still be limited because of the numerous and various approaches to WSC. Through viewing WSC based SOA system from a perspective of mechanistic organization, this paper borrows Divide-and-Conquer (D&C) as the generic strategy to narrow down the problem of effort judgment for the entire SOA implementation to that for individual WSCs ". They present the benefits from a classification matrix that is oriented towards efforts, to set hypotheses, assigning scores using a set of rules. "These effort scores are used to facilitate qualitatively judging different effort between different types of WSC approaches, and eventually construct an effort checklist for WSC approaches. Finally, this effort checklist can be used together with D&C algorithm to realize the qualitative effort judgment for WSC based SOA implementations.

In[63], the authors proposed an extended SOA model for service composition and service dependency. The authors established a dependency demands aware service-oriented architecture (DSOA) to specify dependency aware service interactions, i.e., service publication, discovery, composition and binding. The authors claim that traditional Service Oriented Computing (SOC) focuses on service composition for application development.

In[64], the authors proposed a new service composition mechanism based on peer-to-peer (*P2P*) network. An extended state machine model is shown to identify network model from a service. The model describes a service and its

execution patterns. Three execution patterns (AND, OR and Sequential patterns) are basic service composition constructs of the model. The model serves as a basis for service composition algorithm. Service composition algorithm describes the process of service composition in detail. A graphic model based on one of the execution patterns turns out to be the input of the algorithm.

2.6 Summary :

In this chapter, we gave a survey on OCR, W.S and W.S composition. The objective of this survey is to understand the proposed approach that will be explained in the next section.

Chapter 3

OFF-LINE ARABIC HANDWRITTEN OCR BASED ON WEB SERVICES (OAHOCRWS)

3.1 Overview

This chapter details our work by providing the architecture of the off-line Arabic handwritten OCR based on web services (OAHOCRWS).

3.2 Off-line Arabic handwritten OCR based on web service (OAHOCRWS) overview

In this research we proposed a new idea based on a flexible collaboration of some selected strong complementary approaches and techniques. We mean by flexible, the possibility to customize the recognition rate by adding or reducing the number of approaches and technique that will collaborate together empirically. This flexibility can be achieved by implementing our selected Arabic handwritten OCR (AHOCR) as web services as illustrated in figure (3.1) which considers only the collaboration of two techniques.

In this research we focused on the cooperation of some techniques at the recognition stage, for this purpose we used the ***k*-Nearest Neighbors** (KNN) and **Support Vector Machine** (SVM) techniques because of their good performance and complementarity as we will explain next.

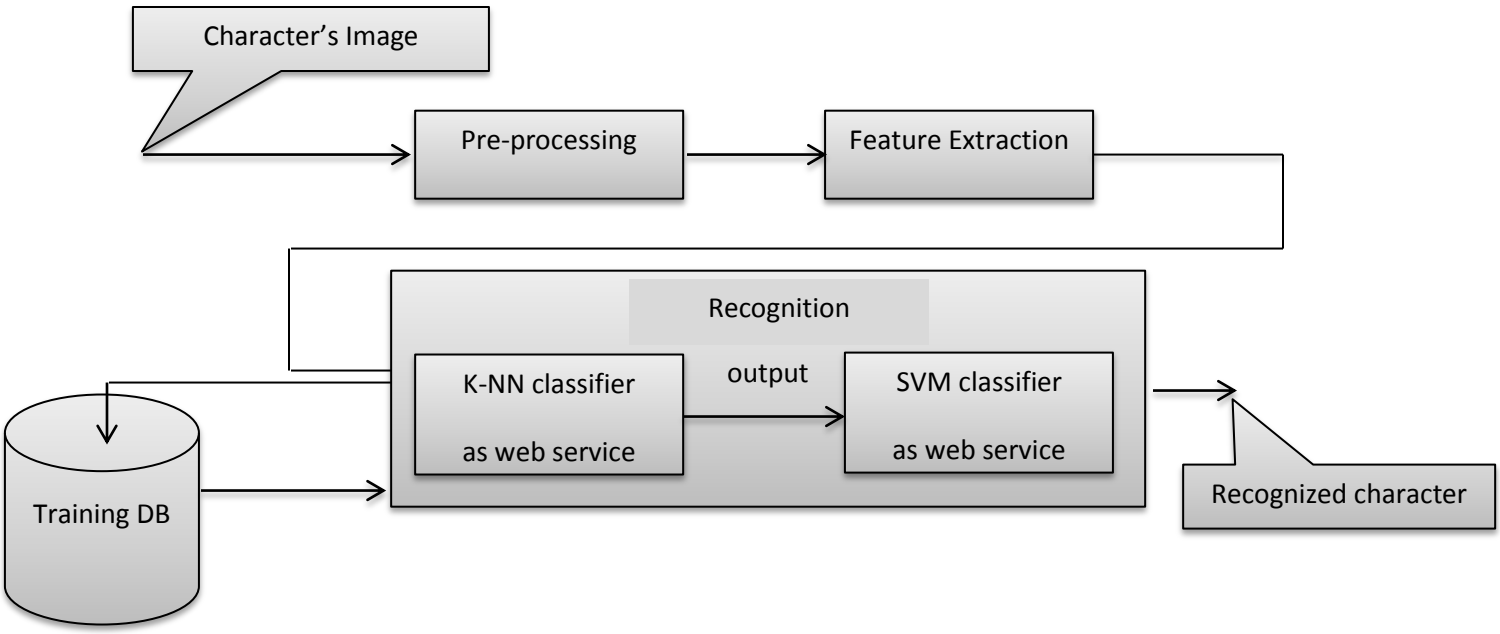


Figure 3.1 : The structure of the proposed Arabic OCR system.

SVM as a decision classifier to exceed the limits of **K-NN** as illustrated in Figure(3.2).

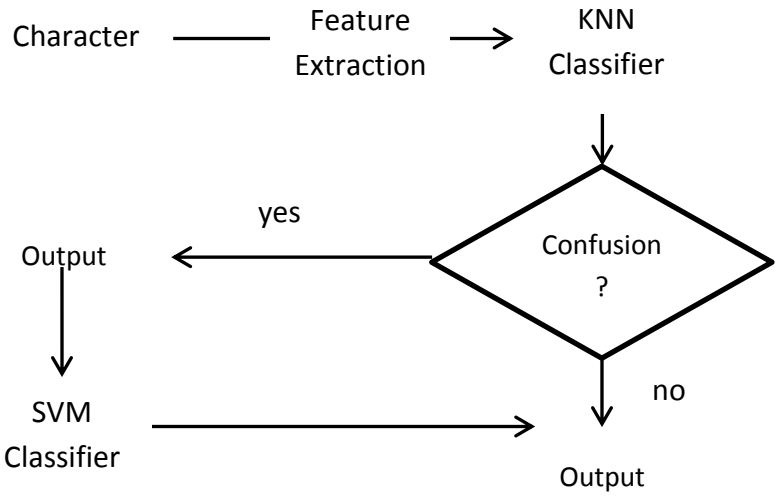


Figure 3.2 : K-NN /SVM mechanism

3.2.1 *k-Nearest Neighbors* (KNN)

KNN is one of the most important non-parametric algorithms for pattern recognition [65]. To classify a character using KNN, the system Determines the k nearest neighbor among the training data sets. and uses the classes of the k nearest neighbors to weight the class candidates[66].

- To determine the class of a new example E:
  - Calculate the distance between E and all examples in the training set : $C_1, C_2, \ldots, C_j$.
  - Select K-nearest examples to E in the training set .
  - Assign E to the most common class among its K-nearest neighbors .

The distance between $E = (E_1, E_2, \ldots, E_j)$ and nearest neighbor C, defined as :

$$D(E,C) = \sqrt{\sum_{i=1}^{n} (E_i - C_i)^2}$$

3.2.2 *Support Vector Machine (SVM)*

SVM is a new type of pattern classifier. This classification technique is based on a novel statistical learning approach[68]. SVM have been applied to handwritten character and digit recognition , also can be applied in different application like face detection, face recognition object detection, object recognition[66]. SVM used to reduce the risk[68].

From many previous studies we have found that the K-NN algorithm achieves good performance for character recognition for different data sets[69,70,71].

K-NN represents a very intersecting classifier for Arabic handwriting recognition because of its great adaptability and versatility in handwriting sequential signals[72].

In this work we have used a hybrid KNN-SVM method for character recognition, Specialized Support Vector Machines (SVMs) are introduced to significantly improve the performance of KNN in handwrite recognition.

When using KNN in the task of handwritten characters recognition, the correct class is almost always one of the K nearest neighbors of the KNN. SVMs are used to detect the correct class among these K different classification hypotheses.

3.2.3 K-NN vs SVM

Table 3.1 : K-NN vs SVM

Technique	**Advantages**	**Disadvantages**
k Nearest Neighbor (k-NN)[73]	1. Training is very fast 2. Simple and easy to learn 3. Robust to noisy training data 4.Effective if training data is large	1. Biased by value of k 2.Computation Complexity 3.Memory limitation 4. Being a supervised learning lazy algorithm i.e. runs slowly 5. Easily fooled by irrelevant attributes
Support vector machines (SVM)[74]	1.Effective in high dimensional spaces. 2. Still effective in cases where number of dimensions is greater than the number of samples. 3. Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient. 4. Versatile: different *Kernel functions* can be specified for the decision function. Common kernels	1. If the number of features is much greater than the number of samples, the method is likely to give poor performances . 2. SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation (see Scores and probabilities, below).

	are provided, but it is also possible to specify custom kernels.	

3.3 Summary

In This chapter we have presented an overview of the general architecture of our proposed technique, in the next chapter will explain the implementation and we report the experimental results.

Chapter 4

Implementation and Performance Evaluation of OAHOCRWS

4.1 Overview

This chapter discusses the implementation details, shows and explains the experiments and tests performed on the (OAHOCRWS) in order to asses its performances.

4.2 OAHOCRWS Implementation

This system consists of three web services and client application, the first web service represents KNN algorithm, the second web service represents the SVM algorithm and the last one represents KNN and SVM algorithms. The three web services and client application have been built using VB.net programming language.

The three web services built in general, so that they can be used from any client application at the same computer or via network.

4.3 The experimental study

To evaluate **OAHOCRWS** system**,** we made various experiments to measure its efficiency. These experiments were conducted on the IFN/ENIT database.

For this purpose we evaluated, in the first hand the importance of the hybrid approach K-NN/SVM in the recognition rate and in the second hand the efficiency of using web services technology on the execution time of the proposed hybrid K-NN/SVM approach, The three web services was created, first for KNN method, second for SVM method, last for KNN and SVM together.

4.3.1 Datasets

Any recognition system needs a large database to train and test the system, one of the these databases is IFN/ENIT which containing 26459 handwritten Tunisian town/village names, 115585 PAWs and 212211 characters scanned at 300dpi and written by 411 different writers.

4.3.2 Experimental environment

The tests were conducted on a local Intel(R) core™ i3-2377 M CPU@1.50 GHz having the following configuration: 4.0 GB of RAM running a Windows7 operating system. VB.NET, visual studio 10 were used to implement and built our OCR application.

4.3.3 Results and discussions

In this section the obtained results are presented. The 28 Arabic characters written with different scripts in different positions in the word representing all the classes used in our experiments. The recognition rate using K-NN and K-NN/SVM is presented respectively in the following table .

Table 4.1 : K-NN and K-NN/SVM recognition rate (%)

Classes	**K-NN Rate(%)**	**K-NN/SVM Rate(%)**
ا	96.20	97.70
ب	95.00	96.51
ت	94.90	96.30
ث	95.40	96.20
ج	96.10	97.30
ح	96.30	97.50
خ	95.30	96.40
د	95.30	96.60
ذ	95.80	96.20
ر	95.40	96.30
ز	95.00	96.30
س	95.15	96.60
ش	95.20	96.25
ص	96.00	96.60
ض	95.92	97.00

ط	96.10	96.80
ظ	96.00	96.95
ع	95.80	96.65
غ	95.95	96.80
ف	96.00	96.80
ق	95.90	97.00
ك	95.80	96.90
ل	96.20	97.10
م	95.60	96.80
ن	95.90	97.00
ه	96.25	97.35
و	96.00	97.10
ي	95.75	96.80
Average(%)	95.70	96.80

The Table(4.1) shows that the hybrid K-NN-SVM classifier improve the performance in terms of recognition rate compared with a single K-NN model for Arabic characters classification process.

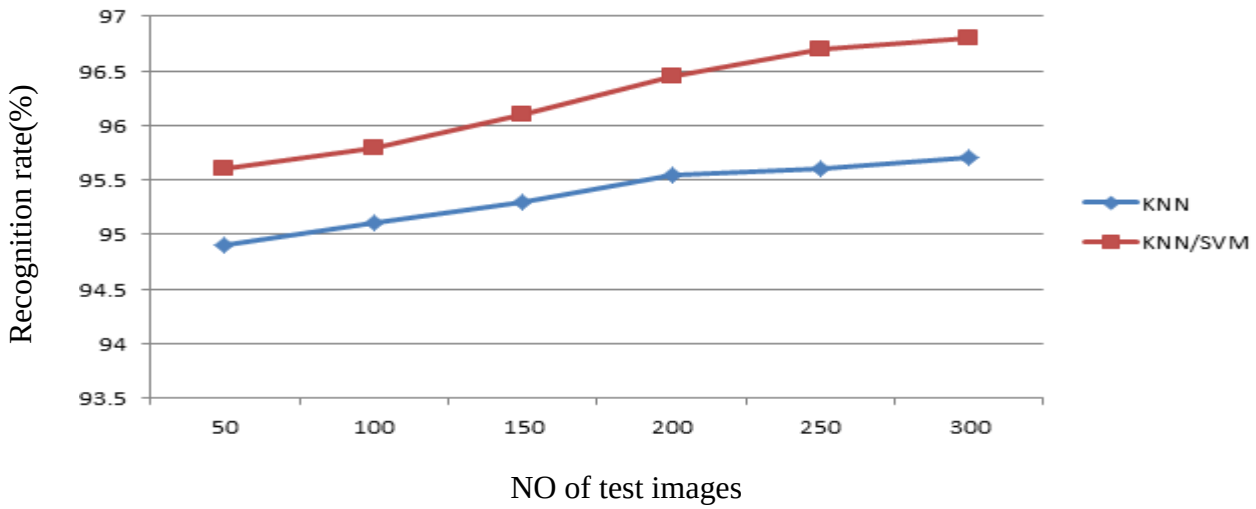


Figure 4.1 : K-NN and K-NN/SVM and their relationship to increase the number of test images

The figure(4.1) shows that the recognition rate of K-NN and K-NN/SVM has increased when the number of test images has increased.

The figure(4.2) below presents the execution time using K-NN and K-NN/SVM with different number of web services.

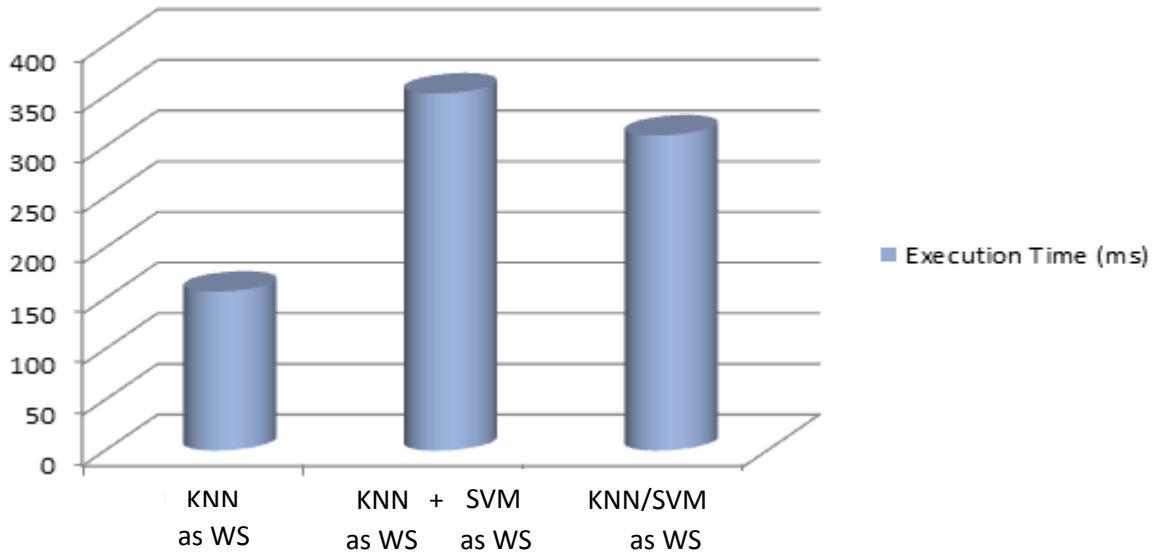


Figure 4.2 **:** Execution time using K-NN and K-NN/SVM with different number of web services

The figure(4.2) shows that the execution time of K-NN is less than the execution time of K-NN/SVM. It shows also that the execution time of K-NN/SVM built as one web service is better than the execution time of K-NN/SVM built as two separated web services.

The figure(4-3) below presents the comparison between K-NN and K-NN/SVM in terms of execution time and recognition rate .

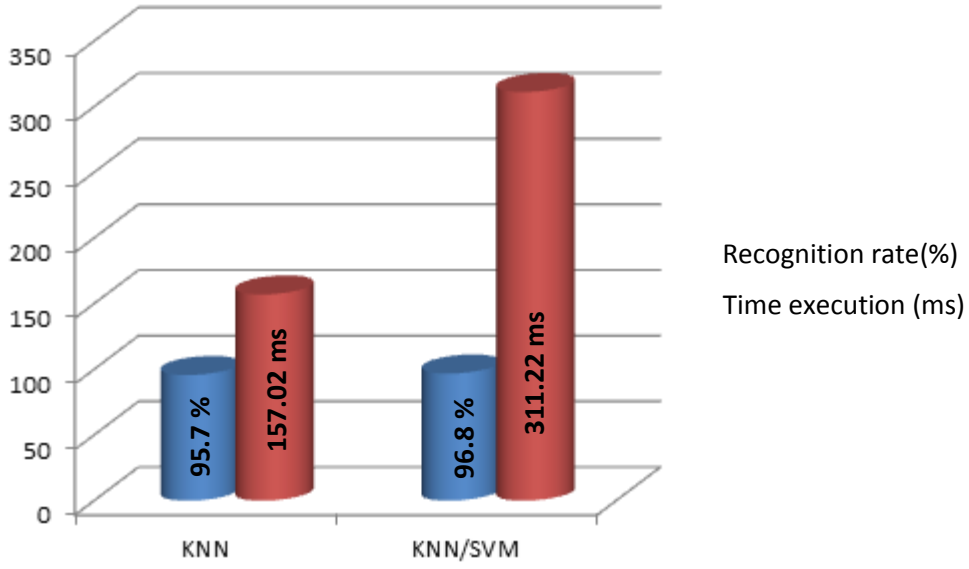


Figure 4.3 : comparison between K-NN and K-NN/SVM

for the execution time and recognition rate

The figure(4-3) shows that the improvement of the recognition rate is unfortunately accompanied with an increasing expense of execution time.

According to the results that we got from the performed tests and experiments, we can say that:

- The hybrid K-NN/SVM classifiers improve the performance in terms of recognition rate compared with single K-NN model for Arabic characters classification process.
- There is a trade-offs between high recognition rate and cost(time), hybrid K-NN/SVM take more time and hence more cost.
- The overall enhancement of recognition rate is about 1.00% compared with single classifier.
- The web service technology can be an appropriate way to make possible the cooperation of several approaches and techniques.

4.4 Summary

In this chapter we discussed the implementation details and performance evaluation of the proposed model , next we conclude this work and we try to give an idea about some future works.

Chapter 5

Conclusion and Future Work

5.1 Overview

This chapter concludes this research work, highlight the corresponding limitations, advantages and finally future works will be presented in order to overcome these limitations.

5.2 Conclusion

In this thesis, an approaches for increasing the recognition rate of Arabic handwritten characters by combining K-NN and SVM using web service technology is proposed. The experiments showed that the combination of K-NN with SVM, have improved performance in terms of recognition rate. The results showed an improvement of 1.00% in recognition rate for Arabic handwritten characters. Also we found that web services technology constitute a good way to make possible the cooperation of several approaches and techniques.

5.3 Future Work

As a future work, we intend conducting more experiments in order to find better complementary approaches that can improve the recognition rate and can lead to build a powerful OCR system.

We intend also to find the optimal way to speed up the execution time of such a powerful OCR system in order to provide users flexible and fast systems.

5.4 Summary

In this chapter we conclude the research ideas and points out the limitations, advantages and the future works that can be done to extend the scope of this research.

LIST OF REFERENCES

[1] **S. Mori, H. Nishida, & H. Yamada**, "*Optical character recognition",* (New York, NY: John Wiley & Sons Inc) Vol. 9, No.2, p. 274-288,(1999).

[2] The tutorials point website(2004) .[Online].Available: http://www.tutorialspoint.com/webservices/what_are_web_services.htm

[3] **R.Saravanan, N. Danapaguiame, K.Rajalakshmi, R.Nivetha, M.Vedhavizhi**, "*An approach for Verifying Web Service composition using Interface Automata* ", International Journal of Scientific & Engineering Research ,Vol.4,NO.2229-5518,(2013) .

[4] **Saeeda Naz , KhizarHayat , MuhammadImranRazzak ,Muhammad Waqas Anwar , SajjadA.Madani , SameeU.Khan , "***The optical character recognition of Urdu-like cursive scripts"*, Vol. 47(3) , Pages 1229–1248 , Pattern Recognition (2013).

[5] **Khorsheed , M.S**, "*Automatic Recognition of Words in Arabic Manuscripts*", Doctor Thesis , Philosophy University of Cambridge ,(2000) .

[6] **A. Amin**, "*Off-line Arabic Character Recognition - the State of the Art", Pattern Recognition*", Vol. **31**, No. **5**, 517-530, (1998).

[7] **F. Al-Fakhri**, "*On-Line Computer Recognition of Hand-Written Arabic Text"*, Master's Thesis,Science University of Malaysia, (1997).

[8] **A. Zeki**, "*Plausable inference Approach to Character Recognition"*, Master's Thesis, National University of Malaysia, (1999).

[9] **Jafaar Alabodi and Xue Li** , "*AN EFFECTIVE APPROACH TO OFFLINE ARABIC HANDWRITING RECOGNITION*" , International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 6, November 2013 .

[10] **M. Amrouch, A. Rachidi, M. Elyassa, D. Mammass**, "*Handwritten Amazigh Character Recognition Based On Hidden Markov Models"*, ICGST-GVIP Journal, Vol.10, Issue 5, pp.11-18, (2011).

[11] **Safabakhsh, R. and P. Adibi** . "*Nastaaligh handwritten word recognition using a continuous-density variable-duration hmm"*. Arabian Journal for Science & Engineering ,vol.30,no (1B), 95—118(2005).

[12] **Jawad et al, "***Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking*", Pattern Recognition Letters, 1081–1088, (2011).

[13] **Alaei, A., P. Nagabhushan, and U. Pal**, "*A Base-line Dependent Approach for Persian Handwritten Character Segmentation"*, Proceedings of the 2010. 20th International Conference on Pattern Recognition, 1977-1980, IEEE(2010).

[14] **N. AZIZI et al, "***OFF-LINE HANDWRITTEN WORD RECOGNITION USING ENSEMBLE OF CLASSIFIER SELECTION AND FEATURES FUSION*", Journal of Theoretical and Applied Information Technology, (2010).

[15] **Sagar. B. M, G. Shobha, and RP. Kumar**, "*OCR for printed Kannada text to Machine editable format using Database approach"*, WSEAS Transactions on Computers, Issue 6, Volume 7, June 2008, ISSN: 1109-2750(2011).

[16] **N. Farah, L. Souici, L. Farah, and M. Sellami** , "*Arabic words recognition with classifiers combination: An application to literal amounts"*. In Artificial Intelligence: Methodology, Systems, and Applications, LNAI 3192, pp. 420-429,(2004) .

[17] **Laslo Dinges et al, "***On-line Automatic Segmentation based Recognition of Handwritten Arabic Words*", (IESK), (2011).

[18] **Nadir, Farah, Ennaji, Abdelatif, KHADIR. Tarek, and Sellami Mokhtar**. "*Benefit of multi-classifier systems for Arabic handwritten words recognition*." In Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on, pp. 222-226. IEEE,(2005).

[19] **Zawaideh, Farah Hanna**. "*Arabic Hand Written Character Recognition Using Modified Multi-Neural Network.*" Journal of Emerging Trends in Computing and Information Sciences (ISSN 2079-8407) 3.7 (2012): 1021-1026.

[20] **Ahmed and Suen**, "*ANOVEL METHOD FOR THE RECOGNITION OF ISOLATED HANDWRITTEN ARABIC CHARACTERS*", Vol. 2,(2014).

[22] **Margner, V. and M. Pechwitz** , "*Synthetic data for Arabic OCR system development"*. In 6th International Conference on Document Analysis and Recognition (ICDAR),vol.3651, pp. 1159—1163,(2003).

[23] **Zaiz Faouzi et al**. "*An Approach Based on Structural Segmentation for the Recognition of Arabic Handwriting*", doi:10.4156/aiss.vol2. issue4.2,(2010).

[24] **Abdelwadood Mesleh et al, "***An Optical Character Recognition*" , Contemporary Engineering Sciences, Vol. 5, no. 11, 521 – 529,(2012).

[25] **Haidar Almohri et al.** "*Real-time DSP-Based Optical Character Recognition System for Isolated Arabic characters using the TI TMS320C6416T*", IAJC-IJME International Conference ISBN 978-1-60643-379-9,(2008).

[26] **Somaya A. S. Al-Ma'adeed** , "*Recognition of Off-line Handwritten Arabic Words*", Doctor Thesis, University of Nottingham , (2004).

[27] **Dutta A K, Chaudhuri S** , "*Bengali alpha-numeric character recognition using curvature features*. *Pattern Recognition",*vol.26(12) , 1757–1770 (1993) .

[28] **G. S. Lehal, C. Singh**, "*A Gurmukhi Script Recognition System* ,Proceedings of 15th International Conference on Pattern Recognition", 2000, Vol. 2, pp. 557-560.

[29] **A. Cheung, M. Bennamoun*, N.W. Bergmann** ,"*An Arabic optical character recognition system using recognition-based segmentation*" , Vol. 34, pp 215-233,(2000).

[30] **Dehghani, A., F. Shabani, and P. Nava** , "*Off-line recognition of isolated Persian handwritten characters using multiple hidden Markov models"*. In International Conference on Information Technology: Coding and Computing vol.28, pp: 506-510 ,(ITCC),(2001).

[31] **Khorsheed, M. S.** , "*Recognizing handwritten Arabic manuscripts using a single hidden Markov model"*. Pattern Recognition Letters vol.24 (14),pp: 2235—2242,(2003).

[32] **Alma'adeed, S., C. Higgins, and D. Elliman** , "*On-line recognition of handwritten Arabic words using multiple hidden Markov models"*. In the Twenty-third SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Volume 17, pp. 75—79, (2004) .

[33] **Souici-Meslati, L. and M. Sellami** , "*A hybrid approach for Arabic literal amounts recognition"*. Arabian Journal for Science and Engineering (AJSE) vol.29,no (2B),pp: 174—194,(2004).

[34] **Al-Hajj, R., C. Mokbel, and L. Likforman-Sulem** , "*Arabic handwriting recognition using base-line dependent features and hidden Markov modeling*" . In :8th international conference on Document Analysis and Recognition (ICDAR) .vol.2,pp. 893-897 (2005).

[35] **N.Farah , A.Ennaji , T.Khader , and M.Sellmi , "***Benefits of multiclassifer systems for Arabic handwritten words recognitions"* . In proc . International conference. On Document Analysis and Recognition .Seoul. Korea , ISBN: 0-7695-2420-6 , pp .222-226(2005).

[36] **Mozaffari, S.,Faez , k.,Ziaratban , M.** , "*Character representation and recognition using quad tree-based fractal encoding scheme"* . In : ICDAR 05 : proceeding of Ninth International conference on Document Analysis and recognition , vol.2. IEEE comput Soc , Seoul (2005) .

[37] **Pechwitz, M. and V. Märgner** . "*HMM based approach for handwritten Arabic word recognition using the IFN/ENIT-database"*. In 7th International Conference on Document Analysis and Recognition (ICDAR),vol.2, pp. 890—894(2003) .

[38] **Pal, U., K. Roy, and F. Kimura** ." *A lexicon driven method for unconstrained Bangla handwritten word recognition"*. In 10th International Workshop on Frontiers in Handwriting Recognition (IWFHR),vol.77, pp. 601—606(2006).

[39**] Menasri, F., N. Vincent, E. Augustin, and M. Cheriet** . "*Shape-based alphabet for off-line Arabic handwriting recognition"*. In 9th International Conference on Document Analysis and Recognition (ICDAR), Volume 2, pp. 969—973(2007).

[40] **Touj, S., N. Ben Amara, and H. Amiri** . "*Arabic handwritten words recognition based on a planar hidden markov mode"l*. International Arab Journal of Information Technology,vol. 2 no(4),(2005).

[41] **Hani Khasawneh** "*A New Algorithm for Arabic Optical Character Recognition*" Proceedings of the 5th WSEAS Int. Conf. on Artificial Intelligence, Knowledge Engineering and Data Bases, Madrid, Spain, February 15-17, 2006 (pp211-224).

[42] **Rajashekararadhya , s.v**," *Efficient zone based feature extraction algorithm for handwritten numeral recognition of popular south Indian scripts*" . Journal of Theoretical and Applied Information Technology vol.7,no(1),1171-1180(2008).

[43] **Dreuw, P., S. Jonas, and H. Ney** . "*White-space models for offline Arabic handwriting recognition"*. In 19th International Conference on Pattern Recognition,pp"1-4, (ICPR)(2008).

[44] **Prem Natarajan, Shirin Saleem, Rohit Prasad, Ehry MacRostie, and Krishna Subramanian, 2008."** *Arabic and Chinese Handwriting Recognition"*, vol. 4768 of Lecture Notes in Computer Science, pp:231–250. Springer, Berlin, Germany(2008).

[45] **Benouareth, A., A. Ennaji, and M. Sellami** . "*Arabic handwritten word recognition using HMMs with explicit state duration"*. Journal on Advances in Signal Processing, Vol. 2008,pp: 1-13 (2005).

[46] **Peeta basu patil and A G Ramakrishnan**. "*Word level multi-script identification*. Pattern recognition letters", Vol.29, pp 1218 – 1229 (2008) .

[47] **S Ben Moussa, A Zahour, A Benabdelhafid, A M Alimi**. "*Fractal – Based system for Arabic/Latin, printed/Handwritten script identification"*.19th International conference on pattern recognition, **ISBN:** 978-142442175-6, pp 1 – 4 (2008).

[48] **Desai, A. A**., "*Gujarati handwritten numeral optical character reorganization through neural network"*, Pattern Recognition,vol.43,no(7),pp 2582 – 2589(2010).

[49] **Pal, U. and A. Sarkar** , "*Recognition of printed Urdu script"*, in Proceedings of the Seventh International Conference on Document Analysis and Recognition (Icdar 2003), Computer Society, Edinburgh, Scotland, pp:1183 -1187(2003).

[50] **Jafaar Alabodi and Xue Li**, "*AN EFFECTIVE APPROACH TO OFFLINE ARABIC HANDWRITING RECOGNITION*" , International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 6, November 2013 .

[51] **Abdurazzag Ali ABURAS and Salem M. A. REHIEL**, " Off-line Omni-style Handwriting Arabic Character Recognition System Based on Wavelet Compression " , ARISER Vol. 3 No. 4123-135 ,2007.

[52] **H. Almuallim and S. Yamaguchi**, "*A method of recognition of Arabic cursive handwriting*", IEEE, Pattern Anal. Mach. Intel.vol. PAMI-9(5), 715 – 722, 715Ð722 (1987).

[53] The Wikipedia website .[Online](2014). Available:http://en.wikipedia.org/wiki/Web_service .

[54] **K.Gottschalk,S.Graham,H. Kreger , J. Snell**,"*introduction to web services architecture*", IBM system journal , Vol 41 , NO 2 , (2002).

[55] W3schools.com, (2014). Web Services - How to Use. [online] Available at: http://www.w3schools.com/webservices/ws_use.asp.

[56] Cavanaugh, (2014) . [online] Altova.com. Available at: http://www.altova.com/whitepapers/webservices.pdf .

[57] **R.Saravanan, N. Danapaguiame, K.Rajalakshmi, R.Nivetha, M.Vedhavizhi** , "*An approach for Verifying Web Service com-position using Interface Automata* " ,

International Journal of Scientific & Engineering Research ,Vol.4,NO.2229-5518,(2013) .

[58] **Xiaochuan Yi, and Kochut, K**." *A CP-nets-based design and verification framework for Web services composition", Proceedings*. 0-7695-2167-3/04 , IEEE International Conference on Web Services, (2004).

[59] **Jinghai Rao, Xiaomeng Su**, "*A Survey of Automated Web Service Composition Methods* ", Springer Journal , Vol. 3387, pp 43-54, (2005) .

[60] Ter Beek, M., Bucchiarone, A. and Gnesi, S(2007). (n.d.). "*Web Service Composition Approaches: From Industrial Standards to Formal Methods"*. 1st ed. [ebook] Pisa: Istituto di Scienza e Tecnologie dell'Informazione. Available at:http://fmt.isti.cnr.it/~mtbeek/TRWS.pdf .

[61] **Maurice H. ter Beek, Antonio Bucchiarone1, and Stefania Gnesi ,** "*A Survey on Service Composition Approaches From Industrial Standards to Formal Methods*", (ISTI-CNR),(2007).

[62] **Zheng Li and Liam O'Brien**, "*A Qualitative Approach to Effort Judgment for Web Service Composition based SOA Implementations*", International Conference on Advanced Information Networking and Applications, pp. 586-593, (2011) .

[63] **Jiehan Zhou, Daniel Pakkala and Juho Perala**,"*Dependency-aware Service Oriented Architecture and Service Composition*", IEEE International Conference on Web Services, (2007).

[64] **Weihua Zhu, Zhihui Du and Suihui Zhu**," *Dynamic Service Composition Based on Peer-to-Peer Network*", Proceedings of the 2 IEEE International Symposium on Services-Oriented System Engineering, 0-7695-2726-4, (2006).

[65**] Belur V. Dasarathy**, "*Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*", Mc Graw-Hill Computer Science Series, IEEE Computer Society Press, Las Alamitos, California, pp. 217-224, 1991.

[66] **H.Hassan and M.Khemakhem** , "*Arabic Islamic Manuscripts Digitization based on Hybrid K-NN/ SVM Approach and Cloud Computing Technologies*" , University of Sfax , Tunisia (2014) .

[67] **B. Boser, I. Guyon, and V. Vapnik**, "*A training algorithm for optimal margin classifiers*", In Proceedings of Fifth Annual Workshop on Computational Learning Theory, New York, vol.2,pp : 396-404,(1992).

[68] **V. V. C. Cortes**. "*Support vector network. Machine Learning*",vol. 20,pp:273–297, 1995.

[69] **Yang Y. and Liu X.**,"*A Re-examination of Text Categorization Methods*". In: Proceedings of 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. 42-49,(1999).

[70] **Joachims T**., "*Text Categorization with Support Vector Machines: Learning with ManyRelevant Features*". In: Proceedings of the European Conference on Machine Learning [C],(1998).

[71] **Li Baoli, Chen Yuzhong, and Yu Shiwen**, "*A Comparative Study on Automatic Categorization Methods for Chinese Search Engine*". In: Proceedings of the Eighth Joint International Computer Conference [C]. Hangzhou: Zhejiang University Press, 117-120,(2002).

[72] **Cleber Zanchettin, Byron Leite Dantas Bezerra and Washington W. Azevedo**," *AKNN-SVM Hybrid Model for Cursive Handwriting Recognition*", WCCI 2012 IEEE World Congress on Computational Intelligence June, 10-15, (2012).

[73] **T. M. Cover and P. E. Hart**, "*Nearest Neighbor Pattern Classification*", IEEE Trans. Inform. Theory, Vol. IT-13, pp 21-27, Jan 1967.

[74] Available at: http://scikit-learn.org/stable/modules/svm.html .