The automatic segmentation and recognition of text on scanned image documents has enabled many applications such as editing of previously printed documents and books, searching for words in that image documents etc. The off-line handwriting segmentation and recognition field arouses great interest in researchers, since there is a high level of ambiguity and complexity in such kind of image documents, and because of the necessity of Optical Character Recognition (OCR) in lots of application especially in office automation. Segmentation and Recognition of cursive handwritten text is the most difficult case in the field of OCR. Much less research has been done on the task of segmentation and recognizing of Arabic texts. The domain of handwriting in Arabic script presents unique technical challenges and has been addressed more recently than other domains due to the necessity of processing tones of previously printed Arabic books and documents. The objective of this thesis is to provide a better way to segment and recognise off-line handwritten Arabic documents.

This chapter describes the concept of OCR and its importance and types. It provides an overview of OCR. This chapter also gives an overview of Arabic language and its characters and a brief history of it. In section 1.6, the research problem of this thesis i.e. segmentation and recognition of off-line Arabic handwritten text problem is clearly defined. The main problem of such kind of OCR applications addressed towards Arabic language come from the cursive nature of Arabic writing. This chapter also summarizes the thesis objective of building a segmentation and recognition of off-line Arabic handwritten documents system. Finally, there is a summary of how this thesis is organized.
1.1: Optical Character Recognition

Optical Character Recognition, usually abbreviated to OCR, refers to the branch of computer science that involves mechanical or electronic conversion of images of printed text or handwritten, usually scanned by a scanner or captured by a camera, into a fully machine-editable text which can be used in text processing applications such as Microsoft Office Word as it had been typed through the keyboard.

The automatic recognition of offline handwritten text could be applied in many areas, for instance ‘form-filling’ applications such as handwritten postal addresses, cheques, credit card sales slips, insurance applications, mail order forms, tax returns etc. Handwritten script, from an unconstrained population of writers and writing, is generated by using OCR applications, must subsequently be processed off-line by computers.

OCR has five major stages as follows:

1. Preprocessing
2. Segmentation
3. Feature extraction
4. Training and Recognition
5. Post Processing

Figure 1.1: Block Diagram of OCR
Chapter 1: Introduction

The printed text is a bit easy for recognition due to its constraint font, whereas recognition of handwritten character is complicated task due to the unconstrained shape variations, different writing style and different kinds of noise that breaks the strokes, primitives in the character or changes their topologies [1].

1.2: OCR: Historical Background

Historically, OCR has evolved in three successive ages. The early one started in 1900 when the Russian scientist Tyuring attempted to help visually handicapped people by developing an aid [2]. In 1929, Gustav Tauschek obtained a patent on OCR in Germany, followed by Handel who obtained a US patent on OCR in USA in 1933. Gustav Tauschek also invented an OCR mechanical device using templates and was granted US patent in 1935. The work done during the first stage was concentrated upon machine-printed text and upon small sets of well handwritten text and symbols. Template matching technique was used to recognize machine-printed text. For handwritten text, features of input text were extracted using primitive feature extraction techniques. These features were fed to statistical classifiers to recognize the input text. In the middle of the 1940s, the first character recognizer was developed [3]. In 1955, the first commercial system was installed at the Reader's Digest, which used OCR to input sales reports into a computer. It converted the typewritten reports into punched cards for input into the computer in the magazine’s subscription department, for help in processing the shipment of 15-20 million books a year [4]. The first Arabic character recognition research started as early as 1975 when Nazif presented his Master’s thesis [5].

Due to lack of computing power, significant works were not performed until the 1980s [6]. The second development stage started in 1980 with the explosion of information technology and availability of computing power. During this stage, structural and statistical approaches were used in many systems [7]-[9].

The first passport scanners were used in the U.S. State Department. These scanners were developed by Caere Corporation in 1984 and some are still in use today. In the coming decades, the use of OCR on passports facilitates the immigration procedures in airports around the world.
Since 1990s, the advanced stage started. Strong methodologies and techniques such as Neural Networks (NNs), Fuzzy set and Support Vector Machine (SVM) were invented to be used in OCR. In addition, computers and other electronic devices such as scanners, cameras and tablets become available in reasonable prices, as well as internet. The final destination of OCR is not reached especially in handwritten cursive text. Furthermore, many excellent OCR systems for Roman based scripts such as Scansoft Omnipage and Abby Fine-Reader are available in the market at reasonable prices.

Lots of efforts have been done on the recognition of Latin, Indian, Chinese and Japanese characters. Unfortunately, little research has been published on the recognition of Arabic characters. This is because of the strongly cursive nature of Arabic writing and diacritic marks. In fact; the techniques applied in other languages are not directly applicable to Arabic characters without fundamental modifications [10].

1.3: Types of Handwriting: Off-line versus On-line

Image documents of handwriting or printed data is converted to digital form either by scanning the image containing the writing or by writing directly with a light pen or any other input device on an electronic surface such as iPad devices etc. These two approaches are distinguished as off-line and on-line handwriting, respectively. Off-line is also called Optical Character Recognition (OCR), without Off-line. The second approach is abbreviated as OLCR i.e. On-Line Character Recognition. Off-line recognition is more difficult than On-line recognition. Plamondon at el. gives a full comprehensive survey of off-line and On-line handwriting recognition [11].

1.4: Arabic Language

1.4.1: Introduction to Arabic Language

Arabic language is the official language for the Arab world, situated in the Middle East and North Africa Arab World consists of 22 countries as shown in figure 1.2. Arabic is a reference to anything connected with Arabia. Mainly, it is a reference to
the Central Semitic language, thus related to and classified alongside other Semitic languages such as Hebrew and Syriac. In terms of speakers, Arabic is the largest member of the Semitic language family. It is spoken by more than 300 million people as a first language and by 300 million more as a second language in some Islamic countries. Standard Arabic is widely taught in schools, universities, and used in the offices and the media [12].

Arabic belongs to the group of Semitic alphabetical scripts in which mainly the consonants are represented in writing, while the markings of vowels (using diacritics) are optional. The earliest-known alphabet to mankind was the North Semitic, which developed around 1700 B.C. in Palestine and Syria. It consisted of 22 consonant letters. The Arabic, Hebrew, and Phoenician alphabets were based on this model. Then, around 1000 B.C., the Phoenician alphabet was itself used as a model by the Greeks, who added letters for vowels. Greek in turn became the model for Etruscan (c. 800 B.C.), whence came the letters of the ancient Roman alphabet, and ultimately all Western alphabets [13]. Figure 1.3 illustrates the root of Arabic language.
Chapter 1: Introduction

1.4.2: Arabic Characters

Arabic alphabet contains 28 letters. Each has between two and four shapes, and the choice of which shape to be used depends on the position of the letter within its word or sub-word. The shapes correspond to the four positions: beginning of a (sub-) word, middle of a(sub-)word, end of a(sub) word and in isolation[14] as shown in Table 1.1.

Arabic script is written from right to left, and letters within a word are normally joined even in machine-print. There is no connection between separate words, so word boundaries are always represented by a space. There is no upper or lower case, but only one case. Letters are connected at the same relative height. The "baseline" is the line at the height at which letters are connected, and it is analogous to the line on which an English word sits [15].

Figure 1.3: The root of Arabic language

Figure 1.4: (1) Individual letters (2) The original word

There are also two small characters called supplementary characters added to some primary characters to represent short vowels. These two supplementary characters are
Hamza and Madah. The positions of Hamza supplementary character are above or below in respect to the primary character. Hamza in some circumstances comes alone such as (علياء) but Madah never. Madah always is found upper the primary character. Table 1.2 shows the supplementary characters and their positions in respect to primary characters.

Table 1.1

Letters of Arabic language and their positions within a word

<table>
<thead>
<tr>
<th>Letter</th>
<th>Isolated</th>
<th>Initial</th>
<th>Medial</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alif</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Baa</td>
<td>بـ</td>
<td>بـ</td>
<td>بـ</td>
<td>بـ</td>
</tr>
<tr>
<td>Taa</td>
<td>تـ</td>
<td>تـ</td>
<td>تـ</td>
<td>تـ</td>
</tr>
<tr>
<td>Thaa</td>
<td>ثـ</td>
<td>ثـ</td>
<td>ثـ</td>
<td>ثـ</td>
</tr>
<tr>
<td>Jiim</td>
<td>جـ</td>
<td>جـ</td>
<td>جـ</td>
<td>جـ</td>
</tr>
<tr>
<td>Haa</td>
<td>حـ</td>
<td>حـ</td>
<td>حـ</td>
<td>حـ</td>
</tr>
<tr>
<td>Khaa</td>
<td>خـ</td>
<td>خـ</td>
<td>خـ</td>
<td>خـ</td>
</tr>
<tr>
<td>Daal</td>
<td>ذـ</td>
<td>ذـ</td>
<td>ذـ</td>
<td>ذـ</td>
</tr>
<tr>
<td>Dhaal</td>
<td>ذـ</td>
<td>ذـ</td>
<td>ذـ</td>
<td>ذـ</td>
</tr>
<tr>
<td>Raa</td>
<td>رـ</td>
<td>رـ</td>
<td>رـ</td>
<td>رـ</td>
</tr>
<tr>
<td>Zaay</td>
<td>زـ</td>
<td>زـ</td>
<td>زـ</td>
<td>زـ</td>
</tr>
<tr>
<td>Siin</td>
<td>سـ</td>
<td>سـ</td>
<td>سـ</td>
<td>سـ</td>
</tr>
<tr>
<td>Shiin</td>
<td>شـ</td>
<td>شـ</td>
<td>شـ</td>
<td>شـ</td>
</tr>
<tr>
<td>Saad</td>
<td>صـ</td>
<td>صـ</td>
<td>صـ</td>
<td>صـ</td>
</tr>
<tr>
<td>Daad</td>
<td>ضـ</td>
<td>ضـ</td>
<td>ضـ</td>
<td>ضـ</td>
</tr>
<tr>
<td>Taa</td>
<td>طـ</td>
<td>طـ</td>
<td>طـ</td>
<td>طـ</td>
</tr>
<tr>
<td>Dhaa</td>
<td>ظـ</td>
<td>ظـ</td>
<td>ظـ</td>
<td>ظـ</td>
</tr>
<tr>
<td>Ayn</td>
<td>عـ</td>
<td>عـ</td>
<td>عـ</td>
<td>عـ</td>
</tr>
<tr>
<td>Ghayn</td>
<td>غـ</td>
<td>غـ</td>
<td>غـ</td>
<td>غـ</td>
</tr>
<tr>
<td>Faa</td>
<td>فـ</td>
<td>فـ</td>
<td>فـ</td>
<td>فـ</td>
</tr>
<tr>
<td>Qaaf</td>
<td>قـ</td>
<td>قـ</td>
<td>قـ</td>
<td>قـ</td>
</tr>
<tr>
<td>Kaaf</td>
<td>كـ</td>
<td>كـ</td>
<td>كـ</td>
<td>كـ</td>
</tr>
<tr>
<td>Laam</td>
<td>لـ</td>
<td>لـ</td>
<td>لـ</td>
<td>لـ</td>
</tr>
<tr>
<td>Miim</td>
<td>مـ</td>
<td>مـ</td>
<td>مـ</td>
<td>مـ</td>
</tr>
<tr>
<td>Noun</td>
<td>نـ</td>
<td>نـ</td>
<td>نـ</td>
<td>نـ</td>
</tr>
<tr>
<td>Haa</td>
<td>هـ</td>
<td>هـ</td>
<td>هـ</td>
<td>هـ</td>
</tr>
<tr>
<td>Waow</td>
<td>وـ</td>
<td>وـ</td>
<td>وـ</td>
<td>وـ</td>
</tr>
<tr>
<td>Yaa</td>
<td>يـ</td>
<td>يـ</td>
<td>يـ</td>
<td>يـ</td>
</tr>
</tbody>
</table>
### Primary and Supplementary characters and their positions in respect to primary characters

<table>
<thead>
<tr>
<th>Primary character</th>
<th>supplementary character</th>
<th>Primary character with supplementary character</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waow</td>
<td>Hamza</td>
<td>ڇ</td>
</tr>
<tr>
<td>Alif</td>
<td>Hamza</td>
<td>₤</td>
</tr>
<tr>
<td>Alif</td>
<td>Madah</td>
<td>⌜</td>
</tr>
<tr>
<td>-</td>
<td>Hamza</td>
<td>⦗</td>
</tr>
</tbody>
</table>

### 1.5: Off-line Handwritten Arabic Text Recognition: A Survey

Earlier surveys of Arabic machine-print and handwriting recognition discussed both machine – print and handwriting, with much more discussion of machine-print, because most of the researches focused on machine-print[16]-[19]. The first researcher in the field of Arabic character recognition is Nazif A. who presented his master’s thesis at Cairo University in 1975 [20]. Due to variations and multi fonts and shapes of Arabic language, in 1980, Nouh et al. [21] suggested a standard Arabic character set to facilitate computer processing of the proposed set. In 1986 Amin and Masini proposed a system for segmentation and recognition Arabic machine-print characters that used horizontal and vertical projections and shaped-based primitives [22]. The system was tested on 100 multi-font words, it achieved a character recognition rate of 85% and a word rate of 95%. In 1988 El-sheikh and Guindi [23] presented recognition system. The positions of segmentation points were based on minimal heights of word contours. The Fourier descriptors are used for classification. In 1990, El-Dabi et al. [24] presented a recognition system for typewritten Arabic. The system used recognition based segmentation technique. Features are extracted using Invariant Moments. The first hand printed Arabic characters recognition system using neural networks was proposed by Amin et al. [25] in 1996. The proposed technique combines rule-based (structural) and classification tests. The technique is more efficient for large complex sets, such as Arabic characters. The feature
extraction in that system is inexpensive. Finally, the execution time is independent of both the character font and size. In 1995 Al-Badr and Haralick [26] presented a holistic recognition system based on shape primitives that were detected with mathematical morphology operations. In 1997 Alherbish et al. [27] presented a first parallel OCR algorithm for Arabic character recognition. In 1999 Khorshed and Clocksin [28] presented a system used features from a word’s skeleton for recognition without prior segmentation. In the same year, BBN presented a system for English and Arabic in which the lexicon was unlimited [29]. The BBN system used The DARPA Arabic OCR Corpus for testing.

In the beginning of the 21 century, Trenkle et al. presented a method that used decision trees for recognition of low-quality and low-resolution images [30]. In 2002 Maddouri and Amiri [31] proposed a recognition system based on combining a global and local vision modelling of the word developed for Latin word recognition by M. Cote. The system has a lot of drawbacks, one of the drawbacks of this system is in its assumption that diacritical marks and dots are always separated, which is not the case with handwritten Arabic. In 2002 Hamami and Berkani [32] developed a system using a structural approach to handle many fonts and sizes. The system includes rules to prevent over-segmentation. In 2002 also Al-Ohali et al. [33] used an HMM to classify handwritten words used in Arabic cheque filling applications. In this system geometrical features were used. In 2003 Amin A. presented a recognition system for hand-printed based on structural description and inductive logic programming [34]. In 2004 Al-Qahtani and Khorsheed presented a system based on the portable Hidden Markov Model Toolkit[35]. In 2004 Somya Alma’adeed presented a recognition system using HMM and multiple HMM [36]. In 2005 Lorigo proposed [37] a novel algorithm for the segmentation and pre-recognition of offline handwritten Arabic text. The character segmentation method over-segments each word, and then removes extra breakpoints using knowledge of letter shapes. This system was tested on a test set of 200 images, 92.3% of the segmentation points were detected correctly, with 5.1% instances of over-segmentation. The pre-recognition component annotates each detected letter with shape information, to be used for recognition in future work. In 2006 Liying Zheng [38] proposed a new method for recognition machine printed Arabic characters. The proposed method employs Ishii et al. chaotic neural network model, which is called globally coupled map using the symmetric map (S-GCM), for
recognizing Arabic characters. Two fonts were used to test the proposed method, Simplified Arabic and Arabic Transparent, with different sizes. The recognition rate is greater than 97%. In 2007 Sarhan and Helalat presented [39] an Arabic character recognition system based on Artificial Neural Networks (ANN) and statistical analysis of the Arabic characters. This system uses binary images. In this system, each typed Arabic character are used as input to a simple feature extraction method, whose output is fed to an ANN that consists of two layers. Simulation results are provided and show that the proposed system always produces a lower Mean Squared Error (MSE) and higher success rates than the current ANN solutions, especially when the contaminating noise level is low.

In 2008 Al-Ma’adeed et al. [40] built a writer identification system using edge-based directional probability distribution features for Arabic words. The researchers studied the feature extraction and recognition operations on Arabic text. To test this system, the researcher built a new database of off-line Arabic handwriting text to be used for writer identification research. The proposed database is meant to provide training and testing sets for Arabic writer identification research.

In 2009 Alkhateeb et al. [41] presented a system of three stages, i.e. preprocessing, feature extraction and classification. Firstly, words are segmented from input scripts and also normalized in size. Secondly, each segmented word is divided into overlapping blocks. Absolute mean values computed for each block of segmented words constitutes a feature vector. Finally, the resulting feature vectors are used to classify the words using the K nearest Neighbor classifier (KNN). The proposed system has been successfully tested on the IFN/ENIT database. Experimental results show a good recognition rate when compared with other methods. In 2010 Jeffrey Woodard et al. used the generative model of computer vision, along with local features represented by quantized Scale Invariant Feature Transform (SIFT) descriptors, to classify writers based on images taken of Arabic text documents. It is the first known application of the method to automated writer recognition. This statistically based approach does not exploit spatial relationships among image features, nor demand explicit segmentation of linguistic units, and does not require supervised training or pre-processing. A performance of 98.0% correct Rank-1
retrieval was achieved for 51 writers, each of whom wrote three cursive samples of the "Rabbit Letter." Although the text of each document in this study was the same, the techniques reported here are designed to be text independent [42]. In 2011 Kanoun at el. proposed a new linguistic-based approach called the affixal approach for Arabic word and text image recognition. The proposed approach uses the linguistic concepts of the vocabulary to direct and simplify the recognition process. The principal contribution of the proposed approach is to be able to categorize the word hypotheses in words that are either derived or not derived from roots and to characterize morphologically each word hypothesis in order to prepare the text hypotheses for later analyses [43]. Leila et al. presented an off-line Multiple Classifier System (MCS) for Arabic handwriting recognition. The MCS combine two individual recognition systems based on Fuzzy ART network used for the first time in Arabic OCR, and Radial Basis Functions. Hu and Zernike Invariant moments are used. For deriving the final decision, different combining schemes are applied. The best combination ensemble has a recognition rate of 90.1%, which is significantly higher than the 84.31% achieved by the best individual classifier.

1.6: Existing Arabic Databases

Unfortunately, there is no standard database for Arabic handwritten recognition. In developing character recognition systems for any language not only Arabic it is necessary to create a database of words. A standard database of images is needed to facilitate research projects in handwritten text as well as printed text recognition.

In 2010 Amin G. Al-hashim et al. [44] presented a new database. It is a comprehensive database of Printed Arabic text for Arabic text recognition research. It consists of scanned images of different forms of Arabic printed text (viz. book chapters, advertisements, magazines, newspapers, and reports) scanned with 200, 300, and 600 dpi resolutions. A total of 6954 pages are scanned. The database may be utilized by Arabic printed text recognition research community. It may be used as a benchmark database.

In 2008 Ashraf AbdelRaouf et al. presented a database for Arabic printed character recognition. This database consists of six million words. It also considers connected
segments or Pieces of Arabic Words (PAWs) as well as Naked Pieces of Arabic Word (NPAWs); PAWS without diacritics [45]. This database is a part of research project for off-line Arabic character recognition systems.

In 2006 The Linguistic Data Consortium (LDC) at the University of Pennsylvania produced a database called “Arabic Gigaword Second Edition” [46]. This is a huge database of 1,500 million Arabic words. The primary drawback of this data base is that the format of the database is in paragraphs and not in single words. So no uses of this database in training phase.

In 2003 Al-Ohali et al. of the centre for pattern recognition and machine intelligence (CENPAMRI) in Montrereal-Canada developed databases of images from 3000 checks provided by a banking corporation. These databases are sub-words, Indian digits, legal amounts (numeric amounts written in words), and courtesy amounts (numeric amounts written with Indian digits). The sub-words database contains 29,498 samples, the Indian digits database 15,175, and the legal and courtesy databases 2,499 each [47].

In 2002 Somaya Alma’adeed et al. [48] presented the Arabic Handwritten Database (AHDB), a new database for off-line Arabic handwriting recognition. The database contains samples from 100 different writers, including words used for numbers and in bank checks. It also contains the most popular words in Arabic writing and free handwriting pages on any topic of the writer’s choosing.

In 2002 M.Pechwitz et al. presented IFN/ENIT database and before this there was no standard database for the field of handwritten Arabic recognition. This database was created by the Institute of Communications Technology (IFN) at Technical University Braunschweig in Germany and the Ecole Nationale d’Ingenieurs de Tunis (ENIT) in Tunisia. It consists of 26,459 images of the 937 names of cities and towns in Tunisia, written by 411 different writers. The database is divided into 4 sets to facilitate the sharing of results among researchers as well as performance comparisons [49].

In the same period The Environmental Research Institute of Michigan-USA (ERIM) [50] has created a database of machine printed Arabic documents. This database
consists of 750 pages collected from Arabic books and magazines. This database contains different text qualities saved in an appropriate file formats. The main drawbacks of this database are; it is small and hard to access.

In 2001 the Euro-Mediterranean project created DIINAR.1 Arabic lexical database. It comprises of 119,693 words.

1.7: Research Problem

Lots of researches have been done in the field of recognizing typed and handwritten Latin, Chinese, and Indian characters. A few researches have been made in the recognition of Arabic characters, mainly due to the cursive nature of Arabic language. Unlike most of the other languages, both typed and hand-written Arabic characters are cursive. Furthermore, Arabic characters can take more shapes than Latin characters.

Another problem is the differences of Arabic fonts; i.e., a certain character in a specific font can be misrecognized as a different character in another font. In this language, some character pairs may be combined together to form another character, that is often called a ligature such as (Lam Alef \( \varepsilon \)). Ligatures unfortunately complicate the segmentation task of any Arabic Optical Character Recognition (OCR) system.

This research deals with the segmentation and recognition of off-line handwritten Arabic text. The problem of Arabic handwritten recognition is a result of many factors, which can be summarized as follows:

1. The research deals with off-line handwritten Recognition not On-line recognition.
2. The research studies cursive handwritten Arabic texts, which totally differ from the machine printed case.
3. The research addresses Arabic writing, which also totally differs from English writing in many ways (right to left for example).
4. Arabic characters can have different heights, which lead to noise cases.
5. Arabic has 28 letters, each of which can be connected in three different ways or being isolated depending on the position. Therefore, each character can have up to 4 different forms depending on its position.

1.8: Objectives of the Study

The main objective of this research is developing OCR system for Arabic language that will be able to transform a handwritten text document written on paper into a digital format that can be manipulated by word processor software. The other objectives of this research are as follows:

4. Building an isolated handwritten Arabic characters database for researchers to test. The isolated characters were collected from several writers.
5. Constructing a feature extraction process.
6. Implementing a segmentation technique that divides any cursive word into characters or sub-words.
7. Implementing a recognition procedure that recognizes characters or sub characters.
8. Training the system.
9. Testing the system.
10. Presenting the results and conclusion and determining the future work.

1.9: Structure of the Thesis

The first chapter describes the concept of OCR and its importance and types. It provides an overview of OCR. This chapter also gives an overview of Arabic language and its characters and a brief history of it. A survey of off-line handwritten Arabic text recognition is presented. The existing Arabic databases also are mentioned in this chapter. The remaining of the thesis is organized as follow: Chapter 2 presents
the proposed techniques and research methodologies adopted. Chapter 3 presents a new database for off-line handwriting Arabic text recognition named Isolated Handwritten Arabic Characters Database (IHACDB). Chapter 4 describes the handwritten Arabic numerals (Hindi digits) and presents the experiments that have been done in recognizing handwritten Arabic numerals. Chapter 5 presents the pre-processing and segmentation stages adopted in this thesis. Chapter 6 describes the feature extraction and recognition stage for off-line handwritten Arabic text. This chapter introduces a new technique i.e. a hybrid feature extraction approach for recognition of off-line handwritten Arabic text based on Support Vector Machine (SVM). Chapter 7 shows the experimental results of this thesis. Conclusion and future research work are summarized in chapter 8.

1.10: Conclusion

Chapter one offers a general introduction of the work. It discusses Optical Character Recognition, types of OCR as well as introduction to Arabic language which the study focuses on developing OCR for off-line handwritten Arabic documents. Arabic language and its characters are mentioned. A survey of off-line handwritten Arabic text recognition and the existing Arabic databases are discussed. The research problem, objectives of the study and structure of the thesis are presented in this chapter.

1.11: References


