J. King Saud Univ., Vol. 12, *Comp. & Info. Sci*., pp. 85-116 (A.H. 1420/2000)

Arabic Character Recognition: Progress and Challenges

Pervez Ahmed and Yousef Al-Ohali
Department of Computer Science, College of Computer & Information Sciences
King Saud University, P.O. Box 51178, Riyadh 11543, Saudi Arabia

(Received 24 February 1998; accepted for publication 16 March 1999)

Abstract. An optical character recognition (OCR) system may provide a solution to the data entry problems, a bottleneck for the data processing industry. Therefore, OCR systems are being developed for almost all major languages and Arabic language is no exception to it. During the past three decades, considerable research and development works have been done towards the development of an efficient Arabic optical character recognition (ACR) system. In this paper we present a comprehensive review of ACR techniques and evaluate the status of the ACR system development and an up to date bibliography.

Keywords: Character recognition; Arabic language; Arabic character recognition; Arabic text recognition; Digital document processing; Pattern recognition.

1. Introduction

Optical character recognition (*OCR*) systems are being developed for almost all major languages and Arabic language is no exception to it [1-199]. Like other languages, Arabic language poses its own challenging problems to the developers of Arabic optical character recognition (*ACR*) systems. During the past three decades, considerable research works have been done towards the development of an efficient *ACR* system. Many research articles and technical reports on this topic have appeared in leading conference proceedings and journals (e.g., [32;157;177]).

The objective of this paper is to review the *ACR* literature and evaluate the status of the *ACR* system development. We believe that the review should prove helpful in identifying and solving problems that are being faced in developing a practical *ACR* system. Although we have compiled a comprehensive bibliography yet this review is based on the research articles that are available to us. In this paper we have developed a set of criteria for categorizing the *ACR* techniques. The criterion set is based on the detailed lists of features definition and extraction, and classification methods that are frequently being experimented with by *ACR* researchers. Besides this, we have tabulated recognition result summary that can be readily used to compare and contrast the performance of respective recognition techniques. Result tabulation has been a problem as some research articles did not provide complete information about the test data that they have used and the environment under which the recognition experiments were conducted.

We have also analyzed characteristics of Arabic text with the recognition point of view. This analysis may give a better understanding of Arabic text and should prove beneficial in assessing the complexities that they may pose in developing an *ACR* system. We present these characteristics in Section 2. General factors that affect *ACR* system development are discussed in Section 3. A set of criteria that we have used to categorize the existing *ACR* techniques is presented in Section 4. Recent advances in *ACR* system design are reviewed in Section 5. Recognition performance issues are discussed in Section 6. The current status of the test data that is being used and its relevance in developing an *ACR* system are examined in Section 7. A conclusion of our study is presented in section 8.

2. Characteristics of Arabic Text

An Arabic text is composed by placing linearly character blocks of varying sizes from right to left. The peculiar characteristic of the Arabic text is that the shape of characters may significantly vary within a word. This variation depends on: the position of the character within a word and its adjacent characters. On the basis of this characteristic we divide Arabic character shapes into two categories: The *primary* category that consists of twenty eight characters (Table 1) and the *secondary* category that consists of eight characters (Table 2). There are four characters that do not fit into this categorization, therefore, they form a special category (Table 3). The characters shown in Table 2 and Table 3 are explained below.

The character Hamzah (ء) causes some confusion. At certain positions it is considered as a sub-component of a character, e.g., Kaf (ك) and at some other positions it is considered as an independent isolated character. The character Alif-Maqhsura (ى) is pronounced as the character Alif while its shape resembles the character Ya (ي). Like the character Alif, it cannot be connected to the succeeding character. The character Ta-marbotah (ة) is pronounced as the character Ta(ت) while its shape resembles the character Ha (ﻫ). Unlike Ta(ت) and Ha(ﻫ), it can't be connected to the succeeding character.

Table 1. Primary category character set

SN	Character name	Shape position			
		End	Middle	Begn.	Isold.
1	Alif	ـا			ا
2	Ba	ـب	ـبـ	بـ	ب
3	Ta	ـت	ـتـ	تـ	ت
4	Tha	ـث	ـثـ	ثـ	ث
5	Jim	ـج	ـجـ	جـ	ج
6	HHA	ـح	ـحـ	حـ	ح
7	KHA	ـخ	ـخـ	خـ	خ
8	Dal	ـد			د
9	Thal	ـذ			ذ
10	Ra	ـر			ر
11	Zai	ـز			ز
12	Sin	ـس	ـسـ	سـ	س
13	Shin	ـش	ـشـ	شـ	ش
14	Sad	ـص	ـصـ	صـ	ص
15	Dhad	ـض	ـضـ	ضـ	ض
16	Tua	ـط	ـطـ	طـ	ط
17	Zua	ـظ	ـظـ	ظـ	ظ
18	Ain	ـع	ـعـ	عـ	ع
19	Gain	ـغ	ـغـ	غـ	غ
20	Fa	ـف	ـفـ	فـ	ف
21	Qaf	ـق	ـقـ	قـ	ق
22	Kaf	ـك	ـكـ	كـ	ك
23	Lam	ـل	ـلـ	لـ	ل
24	Mim	ـم	ـمـ	مـ	م
25	Non	ـن	ـنـ	نـ	ن
26	Ha	ـه	ـهـ	هـ	ه
27	Waw	ـو			و
28	Ya	ـي	ـيـ	يـ	ي

Table 2. Secondary category character set

SN	Character name	Shape position		
		End	**Middle**	**Isolated**
1	Hamzah			ء
2	Alif-Muqsorah	ـى		ى
3	Ta-Marbotah	ـة		ة
4	Hamzah-Waw	ـؤ		ؤ
5	Hamzah-Ya	ـئ	ـئـ	ئ
6	Hamzah-Alif	ـأ		أ
7	Hamzah-Alif	ـإ		إ
8	Alif-Mad	ـآ		آ

Table 3. Special category character set

SN	Character name	Shape position	
		End	**Isolated**
1	Lam-Alif	ـلا	لا
2	Lam-Alif	ـلأ	لأ
3	Lam-Alif	ـلإ	لإ
4	Lam-Alif	ـلآ	لآ

The character Hamzah-waw(ؤ) is pronounced as the character Hamzah(ء) while its main component resembles the character Waw(و). It cannot be connected to succeeding characters. The character Hamzah-ya (ئ) is pronounced as the character Hamzah(ء) while its main component resembles the character Ya(ي). Like Ya(ي), it is connectable to succeeding characters. The characters Hamzah-alif(أ،إ) are pronounced as the character Hamzah(ء) while their main components resemble the character Alif(ا). They cannot be connected to succeeding characters. The main component of the character Alif-mad(آ) resembles the character Alif(ا). It cannot be connected to succeeding characters. Each of the characters Lam-Alif (لا،لأ ،لإ ،لآ), for computer applications (mainly because of the segmentation problem between the Lam(ل) and Alif(ا)), is considered as a new single character in the character set.

In addition to this criteria, the Arabic characters set can also be categorized on the basis of number of constituent components as some Arabic characters are formed by more than one component. Such characters contain one main component and one or more sub-components. The subcomponents are *Noqtah*, *Hamzah* and Mud-sign Table 4.

A *Noqtah* is a dot like sub-component. The count and the position of *Noqtah* relative to the main component form different characters. The maximum count of *Noqtah* in a character is three. The possible combinations of *Noqtah* count and its position relative to the main components in a text are shown in Table 5.

Like *Noqtah,* the appearance of *Hamzah* at a specific position forms different characters. A *Hamzah* may appear: above the characters *Hamzah-waw*(ؤ) and *Hamzah-ya*(ئـ); above or below the characters *Hamzah-alif*(أ،إ) and *Lam-Alif*(لأ،لإ); inside the characters *Kaf*(ك) and *Hamzah-ya*(ئ) or it may appear as a single isolated character (ء). If it is isolated then it is considered as a main component (see Table 6).

Table 4. Special category

Component type	Relative position to main-shape		
	Above	Below	Inside
Dot	خ ض غ ف ن ذ ز	ب جـ	ج ظ ن
Two Dots	ت ق ة	ي	ت
Three Dots	ث ش		ث
Hamzah	أ لأ ؤ ئـ	إ لإ	ك ئ
Mud-sign	آ لآ		

Table 5. Noqtah count

Position	1 dot	2 dots	3 dots
Above the main component	خ ض غ ف ن ذ ز	ت ق ة	ث ش
Below the main component	ب جـ	ي	
Inside the main component	ن ج ظ	ت	ث

Table 6. Characters formed by hamzah

Position	Different Character formation
Above the main component	أ ؤ لأ ئـ
Below the main component	إ لإ
Inside the main component	ك ئ
Single component	ء

The appearance of *Tushkeel* (diacritics) sub-component affects the pronunciation and/or meaning of Arabic words. Its appearance does not form new characters. Different types of diacritics in Arabic text, together with their names are explained below:

Fut'hah: A small horizontal line above character(◌َ); *Kusrah:* A small horizontal line below character(◌ِ) ; *Dummah:* A small comma above character(◌ُ) ; *Tunween:* Double *fut'hah* above character(◌ً), double *dummah* above character(◌ٌ), or double *kusrah* below character(◌ٍ) ; *Shaddah:* A small "w" (◌ّ) above a character. It could be combined with any of the above types. In such a case, the other vowel position will be relative to the *shaddah* instead of the main shape, i.e., a *kusrah* will be under the *shaddah*; *Sokon:* A small circle above the character(◌ْ) ; *Mud-sign* : A small approximation sign appear above the *Alif*(آ) or above *Lam-Alif*(لآ). These two shapes can be seen as another characters in the Arabic language by those who consider the *hamzah-alif* as a different character than *Alif.* Mud-sign can appear above any character in some Arabic fonts (e.g., Othmani font).

Diacritics shapes and corresponding examples are shown in Table 7. An Arabic text may contain punctuation marks and special symbols. Frequently used punctuation marks and special symbols are: a period represented as dot " . "; an excelamination mark represented as " ! "; an interrogation mark represented as " ؟ “; a comma represented as “ ، ”; a semicolon represented as “ ؛ "; an open parentheses represented as ") "; a closed parentheses represented as " ("; Quotation marks represented as " “ " and " ” ".

Table 7. Diacritics shapes with examples

Name	Shape	Example text
Futhah	◌َ	كَتَبَ
Kusrah	◌ِ	كِتاب
Dummah	◌ُ	كُتُب
Tunween	◌ً ، ◌ٌ ، ◌ٍ	كتابٌ
Shaddah	◌ّ ، ◌َّ ، ◌ُّ ، ◌ِّ	كتيِّب
Sokon	◌ْ	يكْتب
Mud-sign	~	قرآن

3. Factors Affecting the Design of Arabic Character Recognition (*ACR*) Systems

We classify factors affecting the development of an *ACR* system into two classes: *random* factors and *linguistic* factors. Random factors affect the document scanning process. Examples of random factors are: the document digitization errors, ink and dirt spattering, paper quality, the quality of writing tools, random distortions introduced by the scanning devices, etc. Linguistic factors are the intrinsic part of the Arabic language. The cardinality of the Arabic alphabet set is one of the linguistic factors that affects the design of *ACR* system. In Arabic character recognition literature, different cardinalities of Arabic alphabet set are considered by the researchers. For example, [67] used 28 basic characters; [108] used 29 basic characters and [157] used 29 basic characters and two additional characters: the Ta-marbotah and the Hamzah. Sometimes the cardinality is increased because different shapes of a character are considered as distinct characters. For example, [157] has considered different shapes of the character Hamzah as different characters as they appear in ء، أ، ؤ، ئ، لأ. For human Arabic text readers this consideration may not make much difference, but for an optical character recognition (OCR) system different representations of such characters need to be considered as different characters because every shape once recognized is mapped into a distinct ASCII code.

Character shape variation within a word is another linguistic factor that affects the design of an *ACR* system. Writing style and character connectivity are the two major sources of shape variations. Shapes of all characters that do not form connectivity with any character remain unchanged in every textual form. Shapes of connectivity forming characters vary drastically (see Table 1). Unlike the development of an OCR system for Latin alphabet based languages, this shape variation poses many problems in developing an *ACR* system for Arabic alphabet based languages.

In addition to characters, an *Arabic* text contains symbols that represent numerals and punctuation marks. Currently used numerals in the Eastern Arab countries are the set of ten Indian digits (See Table 8). In this set, all digits, except the digits 2 and 3, have only one shape form. The digits 2 and 3 are written in two ways in handwritten documents. The decimal point is written as the character *Waw*(,).

Arabic text reproduction method—a linguistic factor— affects design of an *ACR* system. *Arabic* text is written from right to left and it is cursive in nature. Moreover, an *Arabic* word consists of one or more connected components where every component consists of one or more characters. The number of components in a word depends on the number and connectivity property of its characters. *Arabic* words cannot be hyphenated, thus they are written as one unit.

Table 8. The set of Indian digits

Digit as words	Digit as numberal
Zero	٠
One	١
Two	٢
Three	٣
Four	٤
Five	٥
Six	٦
Seven	٧
Eight	٨
Nine	٩

4. Classification of Arabic Character Recognition Techniques

A combination of pattern recognition techniques are being devised to produce an efficient *ACR* system. We classify these techniques into different classes on the basis of the criteria: Textual data acquisition mode; Text style; Text segmentation; Feature definition, extraction and representation; Classification and Post processing.

4.1 Textual data acquisition mode

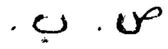
The on-line and off-line are the two possible modes of data acquisition. An on-line *ACR* system recognizes handwritten text by capturing the pen positions in real time.

Research works for developing an efficient on-line character recognition system are described in section 5.1. An off-line *ACR* system recognizes an existing text. In such a system, digital images of existing texts are produced by scanning them line by line or page by page. Afterwards these images are processed and analyzed. An off-line *ACR* system may be built to recognize handwritten and typewritten Arabic texts. Recent advances in the development of an efficient off-line *ACR* system are presented in section 5.2.

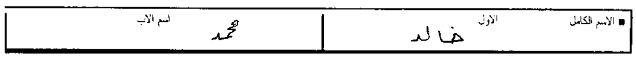
4.2 Text style

An Arabic text can be produced in any of the following styles.

a) Unconstrained non-isolated handwritten text: The natural Arabic text—the cursive handwritten text produced without imposing any constraint on writers. The recognition scheme that recognizes this text requires both the word and character level segmentation (a process that breaks a sentence into isolated words and characters). An example of unconstrained non-isolated handwritten text is shown below:

b) Unconstrained isolated handwritten text: The cursive unconstrained handwritten text produced where writers are expected to write isolated non-overlapping characters like () required for the post office box in a mailing address.

c) Constrained non-isolated handwritten text: The text produced under some constraints like using writing guidelines or writing the text at a fixed position. Filling a printed form is an example of such a text (shown below)



For this type of text, word and character segmentation schemes are required to represent characters to the recognition module.

d) Constrained isolated handwritten text: Isolated characters produced under constrained environment. Filling a form where characters should be written in specified boxes is an example of such text (shown below)

e) Unifont typewritten text: Typewritten text that involves only one font.

f) Multifont typewritten text: Typewritten text involving many fonts.

Each text style mentioned above poses problems of its own and each requires a separate recognition scheme. The recognition schemes developed for solving the problems posed by various text styles are reviewed in section 5.2.1.

4.3 Text segmentation

A text segmentation process extracts the basic constituents from a given text. The basic constituents of an Arabic text are: *a word*, *the complete shape of an isolated character and the partial shape of a character*. In research works where the focus of the research was on the recognition technique only, the segmentation process was ignored. In such works we assume that the test data were manually segmented. However, the segmentation process is an essential step in automating an *ACR* system, therefore, several segmentation approaches were developed. Commonly used approaches for Arabic text segmentation are reviewed in section 5.2.2.

4.4 Feature definition, extraction and representation

A feature describes the characteristics of an underlying character, its partial structure or the structure of the whole word. Like OCR techniques, *ACR* techniques can also be distinguished from one another on the basis of feature definitions that they employ and the way they extract and represent features. In *ACR* research works, both the statistical (quantitative) and structural (qualitative) features are used.

4.5 Classification

In an attempt to obtain a good recognition score, almost all classification techniques: mathematical, statistical, syntactical, graph-theoretic, neural network based, heuristics and so on are used for Arabic character recognition. A comparative performance analysis of these techniques are presented in section 5.2.5.

For some reasons, decision tree (a hierarchical graph-theoretic technique) based classification techniques are popular among *ACR* researchers [4-7; 36; 42; 55; 60-63; 68; 76; 108; 114; 117; 199]. Occasionally, some researchers have referred to this technique as syntactic technique which is a misleading term. Statistical techniques are the second frequent choice [85; 91; 113; 166, 183].

4.6 Post processing

The post processing is a process that uses properties of natural languages to enhance the recognition reliability. Recently, use of well established Markov model based post processing approaches for improving the reliability of an *ACR* system has appeared in *ACR* literature [83].

5. Arabic Character Recognition (*ACR*) System Design: Recent Advances

During the past two decades, attempts to design both the on-line as well as off-line *ACR* systems were intensively made, but as compared to research efforts for devising an off-line *ACR* system, the research efforts for devising an on-line *ACR* system are very little. Major on-line *ACR* techniques developed during the past two decades and their performance are presented in Section 5.1. Off-line *ACR* techniques are reviewed and discussed in Section 5.2.

5.1 On-line *ACR* techniques

The first step in an on-line *ACR* technique is to extract features from the strokes that are formed using stylus as a writing tool in real-time. These features are extracted by segmenting the characters into strokes. Commonly used structural features are: open and closed curve segments, vertical and horizontal strokes, cusp and inflection points, and dot counts [58; 60-63]. However, some researchers have also used hybrid features, i.e., a combination of structural and statistical features. In a system described in [36] the authors have used hybrid features. They formed features by combining the structural feature: direction code with the statistical features: segment length, segment slope, coordinates of each segment point and dot counts.

In all on-line *ACR* techniques reviewed here, a decision tree was used as a classifier without post processing [58; 60]. The salient features of these techniques are summarized in Table 9.

The system described in [58] was tested on 400 characters and 80% correct recognition score was recorded. The system described in [36] uses thirteen different characters. They conducted three experiments. In the first experiment, four words were repeatedly formed using these thirteen characters and used as inputs. The system recognized all the occurrences (390) of these words by recognizing the thirteen training characters at every position. The second experiment was conducted on 50 words (150 characters). In this experiment 86% correct recognition score was recorded. In the third experiment, 50 words of the second experiment were produced by imposing writing constraints. In this experiment 100% correct recognition score was recorded. A summary of the results of these experiments and experiments of other research efforts is given in Table 10.

Table 9. On-line *ACR* systems: summary of recognition techniques

Ref.	Feature type	Classifier	System parameters: Segmentation method	Input data	Recognition type	Comments
[60]	*Qualitative* 1. Open curve 2. Closed curve 3. Vertical strokes 4. Horizontal strokes 5. Cusp points 6. Inflection points 7. Group of dots	A tree Classifier	Stroke Segmentation	Character	Character and word Recognition	It is a real system. It recognizes a word using character recognition.
[61]						Refer to [60]
[58]						Refer to [60]
[36]	Hybrid Features *Qualitative* 1. Direction code 2. Dots *Quantitative* 1. Segment length 2. Segment slope 3. Coordinates of each segment point	A tree Classifier	Stroke segmentation by slope & length measurements	Characters	Character recognition only	It is a simulated system.

Table 10. On-line *ACR* systems: summary of recognition results

Ref.	Training set: Sample size	C%	E%	R%	Test set: Sample size	C%	E%	R%	Over all C%	Speed	Comments
[60]	-	-	-	-	-	-	-	-	-	-	Real system
[61]	-	-	-	-	-	-	-	-	-	-	Refer to [60]
[58]	-	-	-	-	400c	80	-	-	-	-	Refer to [60]
[36]	13c	-	-	-	390c	100	-	-	-	-	Exp1. This experiment was conducted on 390 occurrences of 13 training set characters in four repeated words.
					150c	86	-				Exp2. This experiment was conducted on 50 different words formed by 13 training set characters.
					150c	100	-				Exp3. In this experiment the fifty words of the experiment 2 were produced by imposing some constraints on writers

C: Correct E: Error R: Reject

5.2 Off-line ACR techniques

On the basis of information given in the research articles available to us, we categorize research works on off-line *ACR* systems into four categories. Category I consists of research articles that explicitly state that the technique was developed for off-line *ACR* system. Categories II and III consist of research articles on typewritten and handwritten characters respectively. Although the term off-line was not explicitly mentioned, yet from the text we guess that techniques described in there were intended for the off-line recognition. Category IV consists of research articles where the textual style of the test data was not mentioned at all. Techniques belonging to these categories are summarized in Tables 11-14 respectively. Using the criteria listed in section 4 these techniques are analyzed below.

Table.11. Off-line ACR systems: summary of recognition techniques

Ref	Feature type	Classifier	System parameters				Comments
			Segmentation method	Input data	Recognition type	Hand/ type written	
[198]	*Qualitative* 1. Branch attributes. 2. Loops. 3. Topological description 4. Topological relations *Quantitative* 1. Branch point counts.	Tree structured dictionary of sequence of primitive coding of characters.	Stroke segmenta-tion by skeleton following.	Letters	Word recognition	Hand written	
[117]	*Qualitative* 1. Chain code strokes (primitives) 2. Dot position 3. Connection points 4. End points 5. Junction point 6. Secondary stroke 7. Loop frame 8. Layout context *Quantitative* 1. Dot numbers 2. Relative distance between stroke's starts and end points.	Stagewise classification: • Primary classification involves stroke identification • Secondary classification: combines strokes and recognizes a character. • Postproce-ssing using dictionary lookup.	Stroke segmentation by curve following.	Letters	Character and word recognition	Hand written.	
[197]	*Qualitative* 1. Branch attributes. 2. Loops. 3. Topological description. 4. Topological relations. *Quantitative* 1. Branch point counts.	Tree structured dictionary of sequence of primitive coding of characters.	Stroke segmentation by following direction of writing.	Letters	Word recognition	Hand written.	

Table 12. Typewritten ACR systems: Summary of recognition techniques

Ref.	Feature type	
[89]	*Quantitative* 1. Charater height 2. Character width 3. Number of dots 4. Character area. 5. Pin length—length of a short vertical stroke. 6. Character weight above and below the base line.	*Qualitative* 1. Character position relative to the base line. 2. Dot position 3. Character connectivity 4. Outer boundary shape. 5. Position and value of peaks and vallies in horizontal projection
[85]	*Quantitative* 1. Six moments invariant	
[143]	*Quantitative* 1. Contour segment length 2. Difference between the x and y coordinates of contour segments end points. 3. Angles formed at the intersection point of tangents drawn at two segments end points.	
[68]	*Quantitative* 1. Crossing number 2. Character width	
[30]	*Qualitative* (Morphological features)	
[166]	*Quantitative* Six moments invariant	
[95]	*Quantitative* (Twelve Fourier descriptors)	
[68]	*Quantitative* 1.Horizontal and vertical projection 2.Number of dots.	*Qualitative* Connected component code 1. Dot positions 2. Zigzag shape positions 3. Vertical and horizontal bars 4. North, South, East, West and close curves.
[91]	*Quantitative* (Seven moment invariant descriptors)	
[67]	*Qualitative* (Freeman code)	
[163]	*Quantitative* 1. Qudrants with maximum and minimum daek pixels. 2. Major connectivity directions. 3. Mid point position 4. Single, two and three segment row and columns. 5. Dot count.	*Qualitative* 1.Dot position. 2.Loop positions.
[178]	-	

I. D. = Input Data R. T. = Recognition Type L=Letters A=Alphanumeric N=Numerals C=Character W=Word

Table 12. (Continued) Typewritten ACR systems: Summary of recognition techniques

Ref.	**Classifier**	**System parameters** Segmentation method	I. D.	R. T.	**Comments**
[89]	Classification tree	Histogram based segmentation	L	C	Multi font
[85]	Minimum distance classification	Accumulative Invariant Moments performing: • Line , word, character, segmentation. • Character segmentation is performed during the recognition	L	C	The approach is similar to the approach described in [41]
[143]	Quantitative decision function classifier	Outer contour based segmentation	L	C	
[68]	Classification tree	No	L	C	Isolated characters
[30]	Mathematical Morphology	No	L	C	No practical results
[166]	Minimum distance classification	Accumulative Invariant Moments performing: • Line , word, character, segmentation. • Character segmentation is performed during the recognition	L	C	The approach is similar to the approach described in [60]
[95]	Multi category classification scheme	Outer contour method	A	C	
[68]	Tree classification and Morphological word recognizer	Using connected component codes.	L	CW	
[91]	Minimum distance classifier.	1.Word segmentation by span computing. 2.Character segmentation by recognition.	L	C	
[67]	1.Decision tree for character recognition. 2.Viterbi Algorithm for word recognition.	Segmentation by projection.	L	CW	
[163]	Description matching.	Character segmentation by potential and actual connection column	L	C	Farsi characters recognition
[178]	-	Character segmentation using run length	L	-	Only segmentation

I. D. = Input Data R. T. = Recognition Type L=Letters A=Alphanumeric N=Numerals C=Character W=Word

Table 13. Handwritten ACR systems: Summary of recognition techniques

Ref.	Feature type	Classifier	System parameters: Segmentation method	Input data	Recognition type	Comments
[113]	*Quantitative* 1. Pixel count.	Distance and feature distribution.	No Segmentation	Numerals	Character Recognition	
[114]	*Qualitative* 1. Crossing code	Tree classifier	No Segmentation	Numerals	Character Recognition	Tested with different sizes.
[74]	*Quantitative* 1. Aspect ratio 2. Peak values of projection profile histogram. *Qualitative* 1. Crossing code	Tree classifier.	No Segmentation	Numerals	Character Recognition	
[55]	*Quantitative* Nine moment invariant descriptors.	Linear & Quadratic discriminate function	Character segmentation using histograms	Characters	Character Recognition	Features used Classification rate 1-4 81% 1-6 87.5% 1-7 93% 98.79% Letter position Classification rate Beginning 97.5% Middle 99.17% End 100% Isolated 98.5%
[129]	*Qualitative* • Branch, corner and end point counts. • Dot counts. • Loop count • Aspect ratio. • Character position.	Unspecified	Character segmentation using vertical histograms	Characters	Character Recognition	
[42]	*Qualitative* Strokes	Multistage classifier	Stroke segmentation by curve following	Characters	Character Recognition	
[76]	*Qualitative* Strokes as basic shapes	Tree classifier	Stroke segmentation by contour tracing	Characters	Character Recognition	
[5]	*Qualitative* Crossing point as a basic feature.	Hidden Morkov model and decision-tree classifier	No Segmentation	Numerals	Character Recognition	

Table 14. ACR systems: Summary of recognition techniques

Ref.	Feature type	Classifier	System parameters Segmentation method	Input data	Recognition type	Comments
[140]	*Quantitative* 1. Ten Fourier Descriptors. 2. Curvature Features *Qualitative* 1. Dots and holes.	Minimum distance classifier	No	Letters	Character Recognition	
[83]	-	Neural Network	No	Numerals	Character Recognition	
[70]	-	-	Using Freeman codes	-	-	Only segmentation
[149]	-	Lookup table	-	-	-	

5.2.1 Text style

Research works to build *ACR* systems capable of recognizing text styles ranging from printed and standardized characters to totally unconstrained handwritten text are being carried out. Techniques are devised to recognize handwritten words [197; 198] handwritten characters as well as words [117], handwritten characters [42; 54; 129]; and handwritten numerals [4-7; 74; 113; 114]. Similarly, recognition techniques are being devised for typewritten multifont character recognition [89] and unifont character recognition [67; 68; 85; 91; 108; 143; 164; 166] typewritten alphanumeric recognition [96] and numeral recognition [83]. Some studies emphasize on the development of techniques, hence the text style is not specified [140; 149].

5.2.2 Text segmentation

The text segmentation techniques are explicitly presented by some researchers. However, several researchers, who focus their attention on the development of feature extraction and classification modules only, have limited discussion on segmentation related issues. They assumed that the input to their systems were single isolated characters [4-7; 74; 83; 108; 113; 114; 140]. There are cases where the entire segmentation process is completely ignored [149]. Although text segmentation is a major issue, yet very little effort have been made to study the segmentation problem in isolation [178].

Most of the segmentation processes are described as a preprocessing step. After analyzing the segmentation techniques, we group them into: stroke based segmentation [42; 117; 192; 197] histogram based segmentation [54; 89; 129] outer contour analysis based segmentation [95; 96; 143] connected component based segmentation [68] projection based segmentation [67] potential and actual column connection based segmentation [164] segmentation techniques using run length [178] segmentation using Freeman code [70] word segmentation by span computing [91] and segmentation during recognition [91; 166].

5.2.3 Feature definition, extraction and representation

To obtain an accurate recognition performance, both the quantitative (statistical / numerical) and qualitative (structural / topological) features were defined and used in *ACR* research. Commonly used features are discussed below.

5.2.3.1 Quantitative features

The simplest feature is the black pixel count. In this category there are two features: one is the simple black pixel count in the entire region [113], another is the maximum and minimum pixel counts in a marked region like the quadrants 164]. Other features are: character height [89] character width [89; 108] character area [89], Pin length (the pin length is defined as a short vertical stroke) [89] short line segment count (rows and columns containing a single, two or three pixels long lines) [164]; character weight above and below the base line [89] dot counts (the total number of Noqtah's) [68; 89; 117; 129; 164] aspect ratio (relationship between width and height of a character) [74; 129] projection (peak of projection profile and peak value of x/y histogram of crossing) [74] horizontal and vertical projections [68]; moments and moments invariant [55; 85; 91; 166]. Fourier descriptors [95; 96;148] major connectivity directions [164] contour segment length [143] difference between the x and y coordinates of contour segment end points [129] angle formed by the two tangents drawn at two end points of a segment [143] the relative distance between the start and end points of a stroke[117] chain coded strokes [117].

5.2.3.2 Qualitative features

The qualitative features represent the structure of the entire character or the stroke. Ideally, feature forming structures are assigned a code instead of a value. Examples of the qualitative features are : the branch point count codes [129; 197; 198] branch attributes [197; 198] closed curves [68] open curves [71] corner point count codes [129] crossing code [74; 114] crossing point [4-7]; crossing number: [108]; Freeman code [67] dot positions [68; 89; 117; 163]; end-point codes [117] end-point count code [129] junction points [117] loops [71; 72; 197; 198] loop frame [117;164] loop counts [129], loop position: [164] strokes [42; 144] secondary strokes [117]; topological relations (relations are defined by the nature of endpoint primitives) [197; 198] topological description (combinations of topological relations) [197; 198] layout context (base line information and location of one character with respect to its neighbors) [117] (character position relative to the base line) [89]; character's position within a word (first, last, middle, or isolated character) [129] character connectivity [89]; boundary analysis (outer boundary shape) [89] projection profile (position of peaks and valleys in the horizontal projection) [89] connected component analysis (connected component code) [68] vertical and horizontal bars [68] curves (north, south, east and west facing curves) [68] geometric (direction and curvature features) [194] midpoint position (coordinates of midpoint between uppermost/leftmost and lowermost/rightmost dark points) [164].

5.2.4 Classification

Classification schemes used in *ACR* research are: Tree classifier [4-7; 67; 68; 72; 74; 108; 117; 197; 198], stage-wise classifier [117] that uses three stages: primary, secondary and post-processing. Basic strokes are identified in the primary stage. Using the information of primary stage, characters are recognized in the secondary stage and finally a word is recognized in the post processing stage. [42] uses a multistage classifier: First they classify strokes, and then they combine each string of strokes into a character. Use of the minimum distance classifier is reported in [48; 85; 115; 140; 166]. Use of a statistical decision function is mentioned in [55; 143] uses linear and quadratic discriminate functions. Other classification methods are: mathematical morphology [30; 68] string matching [164]; Viterbi algorithm for word recognition [67]; Hidden Markov model [4-7; 44; 144]; Neural Network [71; 83] and Table lookup [149].

5.2.5 Post processing

After the character recognition process, there might be rejected character(s). In such cases, post processing can be used for further testing. [67] uses word recognition for post processing. They used a dictionary and a probability of observing a given lattice of characters using different models of a word.

6. Performance Analysis

The performance of the on-line *ACR* systems has been analyzed in section 5.1. In this section an analysis of off-line *ACR* systems is presented. Based upon the representation of results in the *ACR* literature, we categorize these test results into four categories described below.

6.1 Off-line character recognition

This category includes those research articles in which the word ‘off-line’ is explicitly mentioned [117; 197; 198]. These articles describe systems that can recognize handwritten words and characters, and printed characters. The performance of the *ACR* techniques reported in these articles is summarized in Table 15.

6.2 Typewritten character recognition

In this category, the research articles that explicitly describe systems for typewritten character recognition are included [30; 67; 68; 85; 89; 91; 95; 96; 108; 143;164; 166]. The performance of these techniques is summarized in Table 16.

Table 15. Off-line character recognition: Result summary

Ref.	Training set S.S.	Training set C%	Training set E%	Training set R%	Testing set S.S.	Testing set C%	Testing set E%	Testing set R%	Over all C%	Comments
[198]	500w	-	-	-	-	-	-	-	86	The test words were produced under the constraints: it must be readable; it must be of fixed width and it must be continuos drawing.
[117]	1200c	92	-	-	-	-	-	-	-	1200 characters were extracted from 200 printed words written in four different types of fonts.
					600c	90				600 characters were extracted from 180 handwritten words in Naskh fonts by three writers. The wriers were asked to rewrite under limited constraints all 200 words of experiment 1.
[197]	500w	-	-	-	-	-	-	-	86	The test words were produced under the constraints: it must be readable; it must be of fixed width and it must be continuos drawing.

S.S.: Sample Size C: Correct E: Error R: Reject

Table 16. Typewritten characters recognition: Result summary

Ref.	Training set S.S.	Training set C%	Training set E%	Training set R%	Testing set S.S	Testing set C%	Testing set E%	Testing set R%	Over all C%	Speed	Comments
[89]	-	-	-	-	-	94	-	-	-	130-240 w/m	Speed depends on the font used.
[85]	900	-	-	-	17	94	-	-	-	10.6 c/m	
[143]	-	-	-	-	4110	99	0.4	0.6	-	-	Laser Printed text
					-	98	1.2	0.8	-	-	A mixture o f 4110 Laser printed and unspecified number of dot matrix printed characters.
[108]	-	-	-	-	-	-	-	-	-	- -	Isolated input characters
[31]	-	-	-	-	-	-	-	-	-	-	No result.
[166]	900w	-	-	-	-	-	-	-	-	-	No result.
[95;96]	-	-	-	-	160w	99	0.5	0.5	-	-	
[68]	50w	-	-	-	1000w	90.04 63.16w 4.87w 2.03w	-	-	-	-	**Low quality text.** 90.04% character recognition yielded 63.1% word recognition. The word recognition rate was improved by 6.9% by applying two post processing operations.
	50w				1000w	95.5 81.6w 6.65w 0.32w					**High quality text.** 95.5% character recognition yielded 81.6% word recognition. The word recognition rate was improved by 6.97% by applying two post processing operations.
[91]	112	100	-	-	-	-	-	-	-	-	Recognition result on original data
	112	98									Recognition result after scaling images by 0.5.
	112	95									Recognition result after scaling images by 0.2.
	112	95									Recognition result after thinning the original data.
	112	87									Recognition result after thinning the scaled images by 0.5
	112	82									Recognition result after thinning the scaled images by 0.2
[67]			-	-	-	-	-	-	90	3 c/s	No experimental detail is given.
[164]	-	-	-	-	-		-	-	100	-	No experimental detail is given
[178]	-	-	-	-	-					-	Only segmentation

S.S.: Sample Size C: Correct E: Error R: Reject

6.3 Handwritten character recognition

The research articles that explicitly describe techniques for handwritten character recognition are the part of this category [4-7; 42; 55; 74; 113; 114; 129]. The performance of these techniques is summarized in Table 17.

Table 17. Handwritten character recognition techniques: Result summary

Ref.	Training set				Testing set				Over all	Speed	Comments
	S.S	C%	E%	R%	S.S	C%	E%	R%	C%	-	
[113]	200	-	-	-	2000c	99	1	-	-		Numeral recognition
[114]	-	-	-	-	500c	99	1	-	-		Numeral recognition. Tested with different sizes and tilts.
[74]	-	-	-	-	150c	99.4	0.6	-	-	-	Numeral recognition.
[55]	-	-	-	-	-	85.599.5	-	-	-	-	Result using linear discriminant function. Result using quadratic discriminant function.
129]	130w	-	-	-	170w	79	3	5	-	-	170 word = about 656 characters 13% enhanced by learning)
						92					Unrecognized due to seg. Errors)
[42]	200w	-	-	-	400w	81.25	2	10	91	-	10% rejection due to multiple choices.
								6.7			6.7 % no decision due to segmentation errors
[5]	-	100	-	-	-	98	-	-	-	-	No sample size given

S.S.: Sample Size C: Correct E: Error R: Reject

6.4 Character Recognition.

In some research articles only the term ‘character recognition’ is mentioned [74; 83; 140; 149]. The recognition results of the techniques described in these articles are summarized in Table 18.

Table 18. Character recognition: Result summary

Reference	Training set				Testing set				Over all C%	Speed	Comments
	S.S.	C%	E%	R%	S.S.	C%	E%	R%			
[140]	500w	99	1	-	5000c	98	2	-	-	-	Estimating number of dots is the major error generator.
[83]	10 10	100 -	-	-	-	-	-	-	-	-	
[70]	-	-	-	-	-	-	-	-	-	-	Only segmentation.
[144]	-	-	-	-	-	-	-	-	-	-	

S.S: Sample Size C: Correct E: Error R: Reject

7. Discussion

The importance of the test data in the development of an OCR system for Latin based languages is very well documented and the same is true in the case of *ACR* system. Unfortunately, there is no standard test data set available that may be used to test and compare *ACR* techniques.

In an attempt to locate a test data, we surveyed available research articles. Our findings are summarized in Tables 19-21 below. From Table 21, it can be seen that the largest test data set consists of 5000 characters only, which is insufficient for an authentic conclusion.

Table 19. Arabic word recognition

Research articles	**Recognition Score** Training data set Sample size	Training data set Correct recog.%	Test data set Sample size	Test data set Correct recog.%	Over all %.	**Test data attributes**
[198]	500	-	-	-	86	Handwritten
[197]	500	-	-	-	86	Handwritten
[129]	130	-	170	79	-	Handwritten
[42]	200	-	400	81.25	-	Handwritten
[166]	900	-	-	-	-	Typewritten
[68]	50	-	1000	63.16	-	Typewritten
			1000	81.60		Typewritten

Table 20. Typewritten arabic character recognition

Research articles	**Recognition score** Training data set Sample size	Training data set Correct recog.%	Test data set Sample size	Test data set Correct recog.%	Over all %.	**Test data attributes**
[89]	-	-	-	94	86	Alpha
[85]	900	-	17	74	-	Alpha
[143]	-	-	4110	98-99	-	Alpha
[108]	-	-	-	-	-	Alpha
[30]	-	-	-	-	-	Alpha
[91]	112	82-100	-	-	-	Alpha
[67]	-	-	-	-	90	Alpha
[164]	-	-	-	-	100	Alpha

Table 21. Handwritten Arabic character recognition

Research articles	**Recognition Score**					**Test data attributes**
	Training data set		Test set		Over all %.	
	Sample size	Correct recog.%	Sample size	Correct recog.%		
[113]	200	-	2000	99.0	-	Numerals
[114]	-	-	500	99.0	-	Numerals
[74]	-	-	150	99.4	-	Numerals
[55]	-	-	-	85.5	-	Numerals
				99.5	-	Numerals
[76]	-	-	-	-	-	Alpha
[6]	-	100.0	-	98	-	Alpha
[114]	500	99.0	5000	98	-	Alpha
[83]	10	100	-	-	-	Alpha
[70]	-	-	-	-	-	Alpha
[149]	-	-	-	-	-	Alpha

The quality and size of the test data set is the only resource which can be used to predict the reliability of an *ACR* technique. A large test data set (more than 1000 character/class) that reflects factors affecting the Arabic text production process is required to estimate the reliability of an *ACR* technique. The common factors that affect the text production process are: font type, pen type, paper texture, paper color, ink color, writing style, writing environment and writers mood.

References

[1] Abbas, S. H., Harba, M. I. and AlMuifraje, M. H., "Optimising the Digital Learning Network for Recognition of the Hand-Written Numerals Used by the Arabs." *Proc. Euro. Conf.,* Paris, France, (April 1986), 505-513.

[2] Abdelazim, H. Y. and Hashish, M. A. "Automatic Recognition of Arabic Text." *Proc. of the 10th Image/ITL Conf.*, IBM Toronto Lab., Toronto, Canada, (Aug. 1987).

[3] Abdelazim, H. Y. and Hashish, M. A. "Arabic Reading Machine." *Proc.of the 10th. National Computer Conference,* King Abdulaziz University, Jeddah, Saudi Arabia, (March 1988), 733-744 .

[4] Abdelazim, H. Y. and Hashish , M. A. "Interactive Font Learning for Arabic OCR." *Proc. of the First Kuwait Computer Conference*, Kuwait, (March 1989), 463-486.

[5] Abdelazim, H. Y. and Hashish, M. A. "Automatic Recognition of Handwritten Hindi Numerals." *Proc. of the CompuEURO '89* : VLSI and Computer Peripherals, Hamburg, Germany, (May 1989), 287-298.

[6] Abdelazim, H. Y and Hashish, M. A. "Automatic Reading of Bilingual Typewritten Text." *Proc. of the CompuEURO'89*: VLSI and Computer Peripherals, Hamburg, Germany, (May 1989), 2/140-144.

[7] Abdelazim, H. Y. *Text Recognition: Theory & Implementation.* Ph.D. Thesis, Cairo University, Dokki, Giza, Egypt. (1989). *

[8] Abdelazim, H. Y., Mousa, A. M., Saleh, Y. L. and Hashish, M. A. "Arabic Text Recognition Using a Partial Observation Approach." *Proceedings of the 12th National Computer Conference*, Riyadh, Saudi Arabia, (Oct. 1990), 427-437.

[9] Abdelazim, H. Y. and Hashish, M. A. "Arabic Typeset: An OCR Approach." *Proc. of 5th Signal Processing Conference* (EUSIPCO' 90) in Barcelona, Spain, (Sep. 1990), 1019-1022.

[10] Abdelazim, H. Y. and Mohammad, A. A.-Maguid "Automatic Reading of Arabic Text with Spell Checking Assistance." *Proceedings of the Conf. on the Use of Arabic Language in Information Technology*, Riyadh, Saudi Arabia, (in Arabic), (May 1992), 1-12.

[11] Abdelazim, H. Y. and Mohammad, A. A.-Maguid "Impact of NLP Models on the Enhancement of Arabic OCR Systems." *Proceedings of the First Int. Conf. on AI Applications*, Cairo, Egypt, (June 1992) II-284/II-291.

[12] Abdelazim, H. Y. and Hybrid, A. "Fuzzy-Neural Approach to the Recognition of Arabic Script." *Proc. of 5th Int. Conf. and Exhibition on Multi-Lingual Computing, Cambridge*, UK, (April 1996), 2.3.1-2.3.13.

[13] Abd El-Gwad, A. O., Salem, M., Abou Shadi, F. and Hesham Arafat, "Automatic Recognition of Handwritten Arabic Characters." *Proc. of the 25th Annual Conf. on Statistics, Computer Science and Operations Research*, Cairo University, Dokki, Giza, Egypt, Vol. 4, (Dec. 1990), 4.63-4.75.

[14] Abdullah, W. H., Saleh, A. O. M. and Morad, A. H. "A Preprocessing Algorithm for Handwritten Character Recognition." *Pattern Recognition Letters*, Vol. 7 (January 1988), 13-18.

[15] Abuhaiba, I.S.I. *Use of Fuzzy Set Theory in Pattern Recognition with Application to Arabic Characters*. M.Phil Thesis, University of Bradford, Bradford, England (1990).

[16] Abuhaiba, I.S.I., Mahmoud, S. A. and Green, R. J. "Cluster Number Estimation and Skeleton Refining Algorithms for Arabic Characters." *The Arabian Journal for Science and Engineering (AJSE),* KFUPM, Dhahran, Saudi Arabia, Vol. 16, No. 4 (1991), 519-530.

[17] Abuhaiba, I.S.I. and Ahmed, P. "Restoration of Temporal Information in Off-Line Arabic Handwriting." *Pattern Recognition*, 26, No. 7 (1993), 1009-1017.

[18] Abuhaiba, I.S.I., Mahmoud, S. A. and Green, R. J. "Recognition of Handwritten Cursive Arabic Characters." *IEEE Transactions on PAMI, Pattern Analysis and Machine Intelligence*, 16, No. 6 (1994), 664-672.

[19] Abuhaiba, I.S.I., Holt, M. J. J. and Datta, S. "Processing of Off-Line Handwritten Text: Polygonal Approximation and Enforcement of Temporal Information." *CVGIP: Graphical Models and Image Processing,* 56, No. 4 (1994), 324-335.

[20] Abuhaiba, I. S. I., Holt, M. J. J. and Datta, S. "Straight Line Approximation and 1D Representation of Off-Line Handwritten Text." *Image and Vision Computing*, 12, No. 10 (1994), 649-659.

[21] Abuhaiba, I.S.I. and Mahmoud, S. A. "Fuzzy Graphs to Recognize Handwritten Arabic Characters." *The Arabian Journal for Science and Engineering (AJSE),* KFUPM, Dhahran, Saudi Arabia, 20, No. 1 (1995), 77-93.

[22] Abuhaiba, I.S.I., Holt, M. J. J. and Datta, S. "Restoration of Temporal Information from Static Images of Handwritten Arabic Script." IEEE SICSPCS '95: *IEEE Singapore International Conf. on Signal Processing, Circuits and Systems '95*, Singapore (July 1995).

[23] Abuhaiba, I.S.I., Datta, S. and Holt, M. J. J. "Line Extraction and Stroke Ordering of Text Pages." *ICDAR'95: Third International Conference on Document Analysis and Recognition,* I, Montreal, Canada (August 1995), 390-393.

[24] Abuhaiba, I.S.I., Datta, S. and Holt, M. J. J. "Processing of Text Documents: Straight Line Approximation and Lost Loop Recovery." *ICDAR'95: Third International Conf. on Document Analysis and Recognition,* Vol. II, Montreal, Canada (August 1995), 1157-1160.

[25] Abuhaiba, I.S.I. Datta, S. and Holt, M. J. J. "Fuzzy State Machines to Recognize Totally Unconstrained Handwritten Strokes." *Image and Vision Computing*, 13, No. 10 (1995), 755-769.

[26] Abuhaiba, I.S.I. Holt, M. J. J. and Datta, S. "Processing of Binary Images of Handwritten Text Documents." *Pattern Recognition*, 29, No. 7 (1996), 1161-1177.

[27] Abuhaiba I.S.I. "Recognition of Off-Line Cursive Handwriting, Computer Vision and Image Understanding." 71, No. 1 (July 1998), 19-38,.

[28] Mohammed Bin Ahmed, Almunji Jo'aly, J. Kraifos and Kaneer, S. "Recognition of Arabic Characters Using Neural Networks for Electronic Document Processing." *Proceedings of the Conf. on the Use of Arabic Language in Information Technology*, Riyadh, Saudi Arabia, (in Arabic) (May 1992), 1-8.

[29] Ahmed P. and Khan, M. A. A. "Computer Recognition of Arabic Script Based Tex t- The State of the Art." *Proc. of the 4th Int. Conf. and Exhibition on Multi-Lingual Computing* (Arabic and Roman Script), Univ. of Cambridge, UK (April 1994), 2.1.1-2.1.10.

[30] Al-Bader, Badr and Haralick, R. "Recognition without Segmentation: Using Mathematical Morphology to Recognize Printed Arabic." *Proc. of the 13th National Computer Conference*, Riyadh, Saudi Arabia (Nov.1992), 813-829.

[31] Al-Badr, Badr. "On the Recognition of Arabic Documents, Technical Report 93-10-01, The Department of Computer Science and Engineering, University of Washington, Seattle, USA (1993).

[32] Al-Badr, Badr and Mahmoud, Sabri A. "Survey and Bibliography of Arabic Optical Text Recognition." *J. of Signal Processing*, 41, No. 1, (Jan. 1995), 49-77.

[33] Al-Bader, Badr and Haralick, R. "Segmentation-Free Recognition of Arabic Text." *Proc. of 5th Int. Conf. and Exhibition on Multi-Lingual Computing,* Cambridge, UK (April 1996), 2.4.1-2.4.13.

[34] AlBaiaty, S. H. and Murad, A. H. "Machine-Printed Arabic Character Recognition Using Template Matching." *Proc. of Second Int. Baghdadon Conf. on Computer Technology & Applications*, Baghdad, Iraq, C4: (1986), 1-6.

[35] Al-Emami, Samir. *Machine Recognition of Handwritten and Typewritten Arabic Characters* Ph.D thesis, Dept. of Cybernetics, University of Reading, Reading, UK (Sep. 1988).

[36] Al-Emami, Samir and Usher, M. "On-Line Recognition of Handwritten Arabic Characters." *IEEE Transactions on PAMI,* Pattern Analysis and Machine Intelligence, 12, No. 7 (July 1990), 704-710.

[37] Al-Fedaghi, Sabah and Amin, Adnan "Automatic Spelling Correction." Technical Report, Electrical and Computer Engineering Department, Kuwait University (in Arabic) (1988).

[38] Ali, Sabah A. *Topological Analysis in the Design of a Machine to Recognise Hand-Printed Characters* Ph.D. Thesis, Brunel University, England, UK (1979).

[39] Ali, Sabah A. and Alsaadoun, Mahdi S. "A Parallel Algorithm for Image Thinning." *Proc. of the First Kuwait Computer Conference*, Kuwait, (in Arabic) (March 1989), 121-140.

[40] Alimi, Adel "An Evolutionary Neuro-Fuzzy Approach to Recognize On-Line Arabic Handwriting." *ICDAR* (1997).

[41] Allam, May "Arabic Character Recognition." *Proceedings of SPIE* - The International Society for Optical Engineering V2181, Conf. On Document Recognition, San Jose, CA: USA, Published by Society of Photo-Optical Instrumentation Engineers, Bellingham, WA, USA. (1994), 351-359.

[42] Almuallim, H. and Yamaguchi, S. "A Method for Recognition of Arabic Cursive Handwriting." *IEEE Transactions on PAMI*, 9, No. 5 (Sep. 1987), 715-722.

[43] Al-Ohali, Yousef and Ahmed, Pervez "A Software Environment for the Development and Evaluation of Arabic Character Recognition System." *Proc. of 5th Int. Conf. and Exhibition on Multi-Lingual Computing*, Cambridge, UK (April 1996), 2.5.1-2.5.10.

[44] Alper, A. and Fatos, Y. "A heuristic Algorithm for Optical Character Recognition of Arabic Script." Signal Processing, 62 (1997), 87-99.

[45] Alqaisy, E.K. and Naser, H. L. "Recognition of Arabic Numerals Using Probabilistic Functions." *Proc. of Computer Processing and Transmission of the Arabic Language Workshop*, Kuwait (April 1985).

[46] Alqaisy, E.K. *Recognition of Hand-Written Arabic Numerals Using Fast Fourier Transform* M.Sc Thesis, National Center for Computers/ Institute of Training and Research, Baghdad, Iraq (in Arabic) (1987).

[47] Alqaisy, E.K. and Naser, H. L. "Using Probabilistic Functions for the Recognition of Handwritten Arabic Numerals." *First Kuwait Computer Conference*, Kuwait (in Arabic) (March 1989), 109-120.

[48] Al-Shebeili, Saleh A., Nabawi, Asim A.-F. and Mahmoud, Sabri A. "Arabic Character Recognition Using 1-D Slices of the Character Spectrum." *Journal of Signal Processing*, 56, No. 1 (Jan. 1997), 59-75.

[49] Al-Tikriti M. N. and Bansal, V. S. "Recognition of Handwritten Arabic Numerals Using Fuzzy Entropy." *J. Eng. Tech*., Baghdad, Iraq, Vol. 2 (1984), 7-20.

[50] Al-Tikriti, M. N. and AlRamahi, S. "A Fuzzy Approach for Some Arabic Handwritten Characters Computer Recognition." *Proc. of Computer Processing and Transmission of Arabic Language Workshop,* Kuwait (April 1985), 1-14.

[51] AlTuwaijri, Majid M. and Bayoumi, Magdy A. "Arabic Text Recognition Using Neural Networks." *Proc. of IEEE International Symp. on Circuits and Systems*, London, England (ABS)Vol. 6 (1994), 415-418.

[52] Al-Yousefi, H. and Udpa, S. “Recognition of Handwritten Arabic Characters.” *Proc. of SPIE, 22nd Annual Technical Symposium on Optical and Opto-Electronics Applied Science and Engineering*, San Diego, CA, USA, Vol. 974 (1988), 330-336.

[53] Al-Yousefi, H. *Recognition of Handwritten Arabic Characters* Ph.D. Thesis, Colorado State University, Fort Collins, USA (Summer 1989).

[54] Al-Yousefi, H. and Udpa, S. “Recognition of Handwritten Arabic Characters via Segmentation.” *Arab Gulf J. Scient. Research*, 8, No. 2 (1990), 49-59.

[55] Al-Yousefi, H. and Udpa, S. “Recognition of Arabic Characters.” *IEEE Transactions on PAMI, Pattern Analysis and Machine Intelligence,* 14, No. 8 (Aug.1992), 853-857.

[56] Adnan, Amin and Azzedine, Kaced. “Reconnaissance des Caracteres Arabes Manuscrits en Temps Reel.” CRIN - Universite de Nancy I Rapport 79- (in French) (1979), 80.

[57] Adnan, Amin Haton, A., Kaced, J. and Mohr, R. “Handwritten Arabic Character Recognition by the IRAC System.” *Proc.5th Int. Joint Conf. on Pattern Recognition*, Miami, Florida, USA (Dec. 1980), 729-731.

[58] Adnan, Amin. “Machine Recognition of Handwritten Arabic Words by the IRAC II System.” *Proc. of 6th Int. Conf. on Pattern Recognition,* Munich, Germany (1982), 34-36.

[59] Adnan Amin and Masini, G. “Machine Recognition of Cursive Arabic Words.” *Proc. of SPIE's 26th Int. Symposium and Instrument Display*, San Diego, California, USA, (Aug. 1982), 286-292.

[60] Adnan Amin, Masini, G. and Haton, J. P. “Recognition of Handwritten Arabic Words and Sentences.” *Proc. of 7th Int. Conference on Pattern Recognition*, Montreal, Canada, (1984), 1055-1057.

[61] Adnan, Amin “State of the Art on Character Recognition, Arabic Language.” Meeting, IBM Europe, Paris, France (Jan.1985).

[62] Adnan Amin, “Arabic Handwriting Recognition and Understanding.” *Proc. of Computer processing and Transmission of the Arabic Language Workshop*, Kuwait, Vol. 1 (April 1985), 1-37.

[63] Adnan Amin, IRAC: Un Systeme pour la Reconnaissance et la Comprehension de l'Arabe Ecrit et Imprime, Thesis of Doctorat d'Etat, CRIN, University of Nancy, France (in French) (1985).

[64] Adnan, Amin and Masini, G. “Machine Recognition of Multi-Font Printed Arabic Texts.” *Proc. of the 8th IEEE Int. Joint Conf. on Pattern Recognition*, Paris: France (Oct. 1986), 392-395.

[65] Adnan, Amin. IRAC: Recognition and Understanding Systems. In: Descout, R. (ed.): *Applied Arabic Linguistics and Signal and Information Processing*. Washington, USA, (1987), 159-170.

[66] Adnan Amin “OCR of Arabic Texts.” *Proc. of the 4th Int. Conf. on Pattern Recognition*, Cambridge, UK, (Mar. 1988), 616-625.

[67] Adnan, Amin and Mari, J. F. “Machine Recognition and Correction of Printed Arabic Text.” *IEEE Trans. on Systems*, *Man and Cybernetics*, 19, No. 5 (Sep. 1989), 1300-1306 .

[68] Adnan Amin and Fedaghi, S. “Machine Recognition of Printed Arabic Text Utilizing Natural Language Morphology.” *Int. J. Man-Machine Studies*, 35 (1991), 769-788.

[69] Adnan Amin “Recognition of Arabic Handprinted Mathematical Formulae.” *Arabian J. For Science and Engineering (AJSE)*, KFUPM, Dhahran, Saudi Arabia, 16, No. 4B (Oct.1991), 531-542.

[70] Adnan Amin and AlSaadoun, H. “A New Segmentation Technique of Arabic Text.*” Proceedings of the 11th IAPR Int. Conf. on Pattern Recognition,* The Hague, Netherlands, (Sept.1992), 441-445.

[71] “Hand-Printed Arabic Character Recognition System Using an Artificial Network”. UK: Pattern Recognition Society, 1996.

[72] Adnan, Amin “Recognition of Printed Text via Machine Learning.” *Proc. ICAPR* (Nov. 1998), 317-326.

[73] Assal, M. A., Kouta, M. M. and Tawfik, B. S. “Recognition of Handwritten Hindu Numbers.” *Proceedings of the 4th ORMA Conf., Military Technical College*, Egypt: Cairo, (June 1992), II-261/II-269.

[74] Assal, M. A. and Fahmy, A. A. “Recognition of Handwritten Hindi Numerals in Low Quality Images.” *Proceedings of the First International Conf. on AI Applications,* Egypt: Cairo, (June 1992), II-261/II-269.

[75] Badie, K. and Shimura, M. “A Classification Method of Arabic Script.” *IECE of Japan*, PRL.78-56 (in Japanese) (1978).

[76] Badie, K. and Shimura, M. “Machine Recognition of Arabic Handprinted Scripts.” *Trans. of IECE, Institute of Electronics and Commun*. Eng. of Japan, E65, No.2 (Feb 1982), 107-114.

[77] Badie, K. and Shimura, M. "Machine Recognition of Arabic Cursive Scripts." In*: Pattern Recognition in Practice.* Gelsema, E.S. and Kanal, L. N. (Eds.), North-Holland Pub. Company, (1982), 315-323.

[78] Bansal, V. and AlTikriti, M. "Recognition of Arabic Handwritten Numerals Using Fuzzy Entropy." *J. for Engineering and Technology,* Baghdad, Iraq, 2, No. 2 (1984), 7-20.

[79] Barber, D. "Automatic Analysis and Tagging of Arabic Words". In: Shivtiel A. (ed.), *Proc. of the Second Leeds Symp.*, England, UK, (1988), 30-35.

[80] Berbar, M. A. *Character Recognition for Arabic Texts* M.Sc. Thesis, Faculty of Electronic Engineering, Menouf, Menoufiyya University, Egypt (Dec. 1994).

[81] Bouhlila, K. "Reconnaissance de Caracteres Arabes." These de Doctorat de Specialite, Ecole Nationnale d'Ingenieurs de Tunis (in French) (1988).

[82] Bouhlila, K. Hamrouni, M. K. and Ellouze, N. "Method of Segmentation of Arabic Text Image into Characters." *First Kuwait Computer Conference*, Kuwait, (March 1989), 442-446.

[83] Braham, Rafik "Applications of a Binary Neural Network." *Proc. of the 13th National Computer Conference,* Riyadh, Saudi Arabia, (Nov. 1992), 900-906.

[84] Dewachi, A. "A Standardization of Arabic Alphabet and Numbers." Internal Report, Iraqi National Computer Center, Baghdad (1975).

[85] El-Dabi, Sharif S., Ramsis, R. and Kamal, A. "Arabic Character Recognition System : A Statistical Approach for Recognizing Cursive Typewritten Text." Pattern Recognition, 23, No. 5 (1990), 485-495.

[86] El-Desouky Ali, Salem Mefreh, El-Gwad, Aida Abd and Arafat, Hesham "A Handwritten Arabic Character Recognition Technique for Machine Reader." *Proc. of the 3rd International Conf. on Software Engineering for Real Time Systems*, Cirencester, UK, (Sep. 1991), 212-216.

[87] El-Desouky Ali, Salem Mefreh, El-Gwad, Aida Abd and Arafat Hesham "A Handwritten Arabic Character Recognition Technique for Machine Reader." *International Journal of Mini and Microocomputers,* 14, No.2 (1992), 57-61.

[88] El-Gowely, Khaled *Recognition of Multi-Font Photo Script Arabic Text*. M.Sc. Thesis, Faculty of Engineering, Cairo University, Dokki, Egypt: Giza, (1989).

[89] El-Gowely, K. El-.Dessouki, O. and Nazif, A. "Multi-phase Recognition of Multi-Font Photoscript Arabic Text." *Proc. of the 10th Int. Conf. on Pattern Recognition*, (1990), 700-702.

[90] El-Hamalaway, Mohamed A. and El-Ramly, Salwa H. "A Novel Arabic Numerals Shapes for Optical Character Recognition." *13th Int. Conf. for Statistics, Computer Science, Social and Demographic Research*, (March 1988), 127-130.

[91] El-Khaly, F. and Sid-Ahmed, M. A. "Machine Recognition of Optically Captured Machine Printed Arabic Text." *Pattern Recognition*, 23, No. 11 (1990),1207-1214.

[92] El-Ramly, S. H. and El-Hamalaway, M. A. "A Language-Dependent Arabic Character Recognition Approach." *14th Int. Conf. for Statistics, Computer Science, Social and Demographic Research*, Vol. 4 (March 1989), 247-254.

[93] El-Ramly, S. H. and El-Hamalaway, M. A. "A New Font for Arabic Characters Simplifies Recognition Procedure." *14th Int. Conf. for Statistics, Computer Science, Social and Demographic Research,* Cairo, Egypt, Vol. 4 (March 1989), 255-261.

[94] El-Ramly, S. H. and El-Hamalaway, M. A. "A New Font for Arabic Characters Simplifies Recognition Procedure." *Proc. of 11th National Computer Conf.,* Dhahran, Saudi Arabia (March 1989), 396-401.

[95] EL-Sheikh, T. S. and Guindi, R. M. "Automatic Recognition of Isolated Arabic Characters." Signal Processing (North-Holland), 14, No. 2 (March 1988), 177-184.

[96] EL-Sheikh, T. S. and Guindi, R. M. "Computer Recognition of Arabic Cursive Scripts." *Pattern Recognition,* 21, No.4 (1988), 293-302.

[97] EL-Sheikh, T. S. and EL-Taweel, S. G. "Real-time Arabic Handwritten Character Recognition." *Third International Conf. on Image Processing and Its Applications*. Warwick, UK, (1989), 212-216.

[98] El-Sheikh, T. S. and El-Taweel, S. G. "Recognition of Typewritten Arabic Characters in Different Fonts." *Proceedings of the IEE Colloquium on Character Recognition and Applications*, England: London, (Oct. 1989), 9/1-9/5.

[99] EL-Sheikh, T. S. and EL-Taweel, S. G. "Real-time Arabic Handwritten Character Recognition." *Pattern Recognition*, 23, No. 12 (1990), 1323-1332.

[100] EL-Sheikh, T. S. and EL-Taweel, S. G. "Segmentation of Handwritten Arabic Words." *Proceedings of the 12th National Computer Conference*, Riyadh, Saudi Arabia, (Oct. 1990), 389-402.

[101] EL-Sheikh, T. S. "Recognition of Handwritten Arabic Formulas." *Proc. of the UK IT 1990 Conference*, Univ. of Southampton, Southampton, England, UK, (March 1990), 344-351.

[102] El-Wakil, Mohamed S. and Shoukry, Amin A. "On-Line Recognition of Handwritten Isolated Arabic Characters." *Proceedings of the First KSU Symp. on Computer Arabization, King Saud Univ*., Riyadh, Saudi Arabia, (April 1987), 109-120.

[103] El-Wakil, Mohamed S. and Shoukry, Amin A. "On-Line Recognition of Handwritten Isolated Arabic Characters." Pattern Recognition, 22, No. 2 (1989), 97-105.

[104] Ashraf Emam, M. AlKhatib, Ismail H. and Korany, E. "Character Recognition of Arabic Script." *Proc. of the 4th Int. Conference and Exhibition on Multi-Lingual Computing* (Arabic and Roman Script), Univ. of Cambridge, UK, (April 1994), 2.1.1-2.1.10.

[105] Mohamed Fakir and Sodeyama, Chuichi "On the Features Extraction for the Recognition of Arabic Scripts." in PRU 91-94, Japan, (December 1991), 49-56.

[406] Mohamed Fakir and Sodeyama, Chuichi "Machine Recognition of Arabic Printed Scripts." SSPR92 (1992).

[107] Mohamed Fakir and Sodeyama, Chuichi "Recognition of Arabic Printed Scripts by Dynamic Programming Matching Method." IEICE Trans. Inf. and Sys., Vol. E76-D, No. 2 (Feb.1993), 235-242.

[108] Fayek, M. B. and Al-Basha, B. "A New Hierarchical Method for Isolated Typewritten Arabic Character Classification and Recognition." *Proc. of the 13th National Computer Conference*, Riyadh, Saudi Arabia, (Nov. 1992), 750-760.

[109] Fehri, M. and Ahmed, M. "A New Approach to Arabic Character Recognition in Multi-Font Document." *Proceedings of the 4th Inter. Conf. and Exhibition on Multi-Lingual Computing (Arabic and Roman Script)*, University of Cambridge, UK, (April 1994), 2.5.1-2.5.7.

[110] Ghuwar, M. and Skarbek, W. "Detection of Baseline for Arabic Text Lines Using Hough Type Transforms." *7th Portuguese Conference on Pattern Recognition, Aveiro, Portugal*, (March 1995), 531-533.

[111] Goneid, A., Salem, A., Tolba, M. and Shaddad, E. "On the Sensitivity of the Recognition Algorithms for Printed Arabic Characters." *Proc. of the 12th Int. Conf. for statistics and Computer Science, Social and Demographic Research*, Cairo, Vol.5 (April 1987), 231-241.

[112] Goneid, A., Salem, A. Tolba, M., and Shaddad, E. "The Sensitivity of Some Recongnition Algorithms for Printed Arabic Characters." *Proc. of the 13th IFIP Conf.*, Tokyo, Japan (1988).

[113] Goneid, A., Mikhail, M. N., Kotkata, H. and Hassan, A. A. "Optical Character Recognition of Arabic Handwritten Numerals." *Proceedings of the 3rd Int. Conference and Exhibition on Multi-Lingual Computing* (Arabic and Roman Script), University of Durham, UK (December 1992), 7.1.1-7.1.10.

[114] Goneid, A., Shalaby, M., Hindawi, S. K., Nazmi, T. and Monir, K. "A Fast Algorithm for the Recognition of Arabic Handwritten Numerals Using Topological Features." *Proceedings of the First International Conf. on AI Applications*, Cairo, Egypt (June 1992), II-253/II-260.

[115] Goneid, A., Mikhail, M. N., Kotkata, H. and Hassan, A. A. "Optical Character Recognition of Arabic Handwritten Numerals." *Proceedings of the First Int. Conf. on AI Applications*, Cairo, Egypt (June 1992), II-270/II-283.

[116] Goraine, H. and Usher, M. "Recognition of Typewritten Arabic Characters in Different Fonts." *Proceedings of the IEE Colloquium on Character Recognition and Applications*. London, England, UK (October1989), 9/1-5.

[117] Goraine, H., Usher, M. and Al-Emami, S. "Off-Line Arabic Character Recognition." *IEEE Computer,* 25, No. 7 (Jul 1992), 71-74.

[118] Guindi, R. M. *An Arabic Text Recognition System* M.Sc. Thesis, Faculty of Engineering, Cairo University, Dokki, Giza, Egypt (1987).

[119] Haj-Hassan, F. "Arabic Character Recognition." *7th Summer Session Arab School of Science and Technology,* Syria, P2.A5-1(1985).

[120] Haj-Hassan, F. *Arabic Character Recognition, Computer and the Arabic Language.* Hemisphere Publishing Corporation, Edited by Pierre Mackay (1990), 113-118.

[121] Haj-Hassan, F. and Haj Ali, W. "Printed Arabic Text Recognition." *The Arabian J. for Science and Engineering* (AJSE), KFUPM, Dhahran, Saudi Arabia, 16, No.4b (Oct.1991), 511-518.

[122] Hashish, M. A., Elkheshen, A. T. and Elghonemy, M. R. "Experiences in Isolated Arabic Word Recognition." *Proceedings of Computer Processing and Transmission of the Arabic Language Workshop*, Kuwait (April 1985), 80.

[123] Hassibi, K. "Machine-Printed Arabic OCR Using Neural Networks." *Proceedings of the 4th International Conference and Exhibition on Multi- lingual Computing* (Arabic and Roman Script), University of Cambridge, UK (April 1994), 2.3.1-2.3.12.

[124] Hyder, S. and Khoujah, A. "Character Recognition of Cursive Scripts." *Proc. of the Regional Conf. on Informatics and Arabization* (IRSIT), Tunis, Vol.1 (June 1988), 118-127.

[125] Hyder, S. and Khoujah, A. "Character Recognition of Cursive Scripts." *Proc. of ACM* (1988), 1146-1150.

[126] Jambi, K and Grace, T. "A New Topological Structural Approach for the Recognition of an Isolated Arabic Word." *Proc. of the Conf. on Information Technology in Support of Economic Development,* Khartoum, Sudan (Dec. 1990).

[127] Jambi, K. *Design and Implementation of a System for Recognizing Arabic Handwritten Words with Learning Ability*. Ph.D. Thesis, Illinois Institute of Technology, Chicago, USA (Aug. 1991).

[128] Jambi, K. "Arabic Character Recognition: Many Approaches and One Decade." *The Arabian J. for Science and Engineering* (AJSE), KFUPM, Dhahran, Saudi Arabia, 16, No. 4B (Oct. 1991), 499-509.

[129] Jambi, K. "A System for Recognizing Handwritten Arabic Words." *Proc. of the 13th National Computer Conference*, Riyadh, Saudi Arabia, (Nov. 1992), 472-482.

[130] Kamel, A., Ramsis, R. R. and El-Dabi, S. S. "Arabic Character Recognition System: a Statistical Approach for Recognizing Cursive Typewritten Text." *Pattern Recognition*, 23, No. 5 (1990), 485-495.

[131] Khella, Fakhry *Analysis of Hexagonally Sampled Images with Application to Arabic Cursive Text Recognition*. M.Phil. Thesis, University of Bradford, Bradford, England, UK (1992).

[132] Khella, Fakhry and Mahmoud, Sabri "Recognition of Hexagonally Sampled Arabic Characters." *Arabian J. for Science and Engineering* (AJSE), KFUPM, Dhahran, Saudi Arabia, 19, No. 4A (Oct.1994), 565-586.

[133] Khemakhem, M. and Fehri, M. C. "Arabic Typewritten Character Recognition Using Dynamic Comparison." *Proc. of the First Kuwait Computer Conference*, Kuwait, (March 1989), 447-462.

[134] Kurdy, Bassam M. and Joukhadar, Ammar "Multifont Recognition System for Arabic Characters." *Proceedings of 3rd International Conference and Exhibition on Multi-Lingual Computing* (Arabic and Roman Script), University of Durham, UK (Dec.1992), 7.3.1-7.3.9.

[135] LaPre, Christopher; Zhao, Ying; Raphael, Christopher; Schwartz Richard; and Makhoul, John "Multi-Font Recognition of Printed Arabic Using the BBN Byblos Speech Recognition System." *Proceedings of the 1996 IEEE International Conf. on Acoustics, Speech and Signal Proc.,* ICASSP. Atlanta, GA, USA, (ABS), Vol. 4 of 6 (1996), 2136-2139.

[136] Mahdi, Hani M. K. "Small Size Image Transformations and Representations." *Proceedings of the 12th Int. Conf. for Statistics, Computer Science, Social and Demographic Research,* Cairo, Egypt, Vol. 5 (1987), 1-14.

[137] Mahdi, Hani M. K. "Thinning and Transforming the Segmented Arabic Characters into Meshes of Unified Small Size." *Proceedings of the 14th Int. Conf. for Statistics, Computer Science, Social and Demographic Research*, Cairo, Egypt, (March 1989), 263-276.

[138] Hossam El-Din Mahgoub and Hazem Y. Abdelazim, Arabic Spelling Verification and Assistance, *Proceedings of the Conf. on the Use of Arabic Language in Information Technology*, Riyadh, Saudi Arabia, (May 1992).

[139] Mahmoud, S. A., AbuHaiba, I.S.I. and Green, R. J. "Skeletonization of Arabic Characters Using Clustering Based Skeletonization Algorithm (CBSA)." *Pattern Recognition*, 24, No. 5 (1991), 453-464.

[140] Mahmoud, S. A. "Arabic Character Recognition Using Fourier Descriptors and Character Contour Encoding, Pattern Recognition." 27, No. 6 (1994), 815-824.

[141] Mahmoud, M. Y., Elhamalaway, M. A. and Fahmy, A. A. "A Statistical Approach for Arabic Character Recognition." *Proceedings of the 12th International Conference for Statistics and Computer Science*, Cairo, Egypt, Vol. 5 (April 1987), 13-3-1:13-3-7.

[142] John Makhoul, C. LaPre, Raphael, C., Schwartz, R. and Ying Zhao, "Towards Language Independent Character Recognition Using Speech Recognition Methods." *Proc. of 5th Int. Conf. and Exhibition on Multi-Lingual Computing*, Cambridge, UK (April 1996), 2.6.1-2.6.6.

[143] Volker Margner and Braunschweig, TU "SARAT- A System for the Recognition of Arabic Printed Text." *Proceedings of IAPR Int. Conference on PatternRecognition*, The Hague, Netherlands, (Sep. 1992), 561-564.

[144] Miled, H., Cheiriet, M. and Olivier, C. "Multi-level Arabic Handwritten Word Recognition." *Proc. Advances in Pattern Recognition (SSPR & SPR),* Sydney, Australia, (Aug 1998), 944- 951.

[145] Murad, A. H. *Recognition of Arabic Characters Using Microcomputers.* M.Sc. Thesis, Univ. of Technology, Baghdad, Iraq (1984).

[146] Nabawi, Asim A-F. and Mahmoud, Sabri A. "Automatic Arabic Character Recognition Using Back-Propagation Neural Networks." *Jounral of King Saud Univ., Computer and Information Sci.* Riyadh, Saudi Arabia, Vol. 9 (in Arabic) (1997), 1-29.

[147] Namane, A. and Sid-Ahmed, M. A. "Character Scaling by Contour Method." *IEEE Trans. on PAMI, Pattern Analysis and Machine Intelligence*, 12, No. 6 (June 1990), 600-606.

[148] Nazif, A. M. *A System for the Recognition of Printed Arabic Characters*. M. Sc. Thesis, Cairo University, Dokki, Giza, Egypt, (1975).

[149] Nouh, A., Sultan, A. and Tolba, R. "An Approach for Arabic Characters Recognition." *J. Eng. Sci., Riyadh Univ.* (King Saud Univ.), Riyadh, Saudi Arabia, Vol. 6 (2), (1980), 185-191.

[150] Nouh, A., Sultan, A. and Tolba, R. "On Feature Extraction and Selection for Arabic Characters Recognition." *Arabic Gulf Journal of Scientific Research*, 2, No. 1 (1984), 329-347.

[151] Nouh, A., Nurul-Ula, A. and Sharaf Eldin, A. "Sequential Processing of Binary Images." *J. of Engineering Sciences*, King Saud University, Riyadh, Saudi Arabia, 13, No. 2 (1987), 229-243.

[152] Nouh, A., Nurul-Ula, A. and Sharaf Eldin, "A. Boolean Recognition Technique for Typewritten Arabic Character Set." *Proceedings of the First KSU Symp. on Computer Arabization*, King Saud Univ., Riyadh, Saudi Arabia, (April 1987), 90-97.

[153] Nouh, A., Nurul-Ula, A. and Sharaf Eldin, A. "The Arabic Character Database System, ACDABAS." *Proc. of the First KSU Symposuim on Computer Arabization*, King Saud Univ., Riyadh, Saudi Arabia, (April 1987), 98-108.

[154] Nouh, A., Nurul-Ula,. A. and Sharaf Eldin, A. "An OCR Based Address Reader for Arabic Postal Systems." *First International Conference of the Regional Institute for Informatics and Telecommunications*, Tunis (March 1988).

[155] Nouh, A., Nurul-Ula, A. and Sharaf Eldin, A. "Algorithms for Feature Extraction: A Case Study for the Arabic Character Recognition." *Proceeding of the 10th. National Computer Conf*., King Abdulaziz University, Jeddah, Saudi Arabia, (March 1988), 653-666.

[156] Nouh, A., Nurul-Ula, A. and Sharaf Eldin, A. "A Proposed Algorithm for Thinning Binary Arabic Character Patterns." *First Kuwait Computer Conference*, Kuwait, (March 1989), 426-441.

[157] Nouh, A. "Recognition of Constrained Handprinted Arabic Characters." *Second International Conference on Arabic Language & The Advances in Information Technology*, Tunis, (1993), 200-214.

[158] Nurul-Ula, A. and Nouh, A. "Automatic Recognition of Arabic Characters Using Logic Statements - Part I: System Description and Preprocessing." *J. Eng. Science*, King Saud Univ., Riyadh, Saudi Arabia, 14, No. 2 (1988), 343-353.

[159] Nurul-Ula, A. and Nouh, A. "Automatic Recognition of Arabic Characters Using Logic Statements - Part II: Developement of Recognition Algorithm." *J. Eng. Science, King Saud Univ*., Riyadh, Saudi Arabia, 14, No. 2 (1988), 355-367.

[160] Olivier C., Miled H. and Lecourtier Y. "Segmentation and Coding of Arabic Handwritten Words." *Proc. 13th ICPR*, Vienna, Austria, (Aug 1996), 264-268.

[161] Parhami, B. and Mavaddat, F. "Computers and the Farsi Language: A Survey of Problem Areas." *Information Processing'77 IFIP Congress*, North-Holland, Amsterdam: Holland, (1977), 673-676.

[162] Parhami, B. "Impact of Farsi Language on Computing in Iran." Mideast Computing, No. 1 (Sep. 1978), 6-7.

[163] Parhami, B. “On the Use of Farsi and Arabic Languages in Computer-Based Information Systems.” *Proc. of Symposium on Linguistic Applications of Computer-Based Information Systems*, New Delhi, India, Session 3, Paper 11 (Nov. 1978).

[164] Parhami B. and Taraghi, M. “Automatic Recognition of Printed Farsi Texts.” *Pattern Recognition*, 14, No.1-6 (1981), 395-403.

[165] Rahho, G., Saadallah, S. and Yacu, S. G. “Design of Bilingual Handprinted Reading Machine Using a Microcomputer.” *Proc. of Second Int. Baghadad Conf. on Computer Technology and Applications,* Baghdad, Iraq, 2 , No. 5 (1986).

[166] Ramsis, R., El-Dabi, S. S. and Kamel, A. “Arabic Character Recognition System.” IBM Kuwait Scientific Center, Kuwait, Report No. KSC027 (Jan. 1988).

[167] Rashid, Mosfeq. “Arabic Character Recognition Using Integrated Segmentation and Recognition.” *Proc. of 5th Int. Conf. and Exhibition on Multi-Lingual Computing*, Cambridge, UK, (April 1996), 2.1.1-2.1.10.

[168] Saadallah, S. and Yacu, S. G. “Design of an Arabic Character Reading Machine.” *Proc. of Computer Processing & Transmission of the Arabic Language Workshop*, Kuwait, (April 1985), 17.

[169] Saadallah, S. and Yacu, S. G. “Design of Bilingual Handprinted Reading Machine Using Microcomputer.” *Proc. of the Second Int. Baghdad Conf. on Computer Technology & Applications*, Baghdad, Iraq, PG5: (1986), 1-6.

[170] Sabri, M. “Recognition of Printed Arabic Characters Using the Architectural Correspondence Technique.” *Proc. of the Second International Baghdad Conf. on Computer Technology and Applications*, Baghdad, Iraq, Vol. I, C.4-1 (1986).

[171] Hermineh, Y. Y. Sanossian, “Arabic Character Recognition System Using Neural Network.” *Proc. of the IEEE Signal Proc. Society Workshop: Neural Networks for Signal Proc*., Kyota, Japan, (ABS), (1996), 340-348.

[172] Seddik, M. and Shoukry, A. “On-Line Recognition of Handwritten Isolated Arabic Characters.” Pattern Recognition, 22, No. 2 (1989), 97-105.

[173] Seflan, A. M. “The Need for a Standard Arabic Character Set.” *Proc. of the Second National Computer Conference,* Riyadh, Saudi Arabia (1975).

[174] Shaddad, E. M., Sharkawy, M. and Tolba, M. “Fourier Descriptors for Printed Arabic Character Recognition.” *Proc. of the 13th Int. Conf. for Statistics and Computer Science, Social and Demographic Research*, Cairo, Vol.5 (1988), 137-146.

[175] Shaheen, Samir I. and Abo-Samra, Gibrael A. “On-Line Arabic Cursive Script Recognition.” *The Egyptian Computer J., ISSR*, Cairo Univ., Dokki, Giza, Egypt, 18, No. 2 (1990), 222-240.

[176] Shoukry, A. and Amin, Adnan “Machine Recognition of Cursive Arabic Words.” In: Tescher, G. (Ed.): *Application of Digital Image Processing IV*, SPIE (1983).

[177] Shoukry, A. “Arabic Character Recognition: State of the Art.*” 11th Saudi National Computer Conference*, Dhahran, Saudi Arabia, (1989), 382-390.

[178] Shoukry, A. “A Sequential Algorithm for the Segmentation of Typewritten Arabic Digitized Text.” *The Arabian Journal for Science and Engineering (AJSE),* KFUPM, Dhahran, Saudi Arabia,Vol. 16, No. 4 (Oct. 1991), 543-556.

[179] Taraghi, M. *Automatic Recognition of Printed Farsi Texts*. M.Sc. Thesis, Deptartment of Mathematics and Computer Science, Sharif University of Technology, Iran (1978).

[180] Tawfik, B., Kouta, M. and Assal, M. “Recognition of Handwritten Hindi Numerals.” *Proceedings of 3rd ORMA Conf., Military Technical College*, Cairo, Egypt (1991).

[181] Techonova, Irina “Arabic Optical Character Recognition.” *Proc. of 3rd Int. Conference and Exhibition on Multi-Lingual Computing* (Arabic and Roman Script), University of Durham, UK, (Dec. 1992), 7.2.1-7.2.7.

[182] Tolba, M. F., A.-Wahab, S. and Salem, A. “A Recognition Algorithm for Arabic Printed Characters.” *11th Int. Conf. for Statistics, Computer Science, Social and Demographic Research*, Cairo, Vo1. 2 (1986).

[183] Tolba, M. F., Goneid, A. Salem, A. and Shaddad, E. “The Sensitivity of the Recognition Algorithm for Printed Arabic Characters.” *12th International Conference for Statistics, Computer Science, Social and Demographic Research*, Cairo, Vo1.5 (April 1987), 231-241.

[184] Tolba,M. F., A.-Wahab, S. and Salem, A. "A Recognition Algorithm for Arabic Printed Characters." IASTED: *International Symposium on Applied Informatics*, Grindelwald, Switzerland, (1987), 128-131.

[185] Tolba, M. F. and Mahdi, H. M. K. "On the Transformation of Arabic Characters into Unified Mesh Size.", *Proc. of the 12th International Conf. for Statistics, Computer Science, Social and Demographic Research*, Cairo, Egypt, (April 1987), 129-136.

[186] Tolba, M. F., A.-Wahab, S. and Salem, A. "A Recognition Algorithm for Printed Arabic Characters." *Proc. of the First KSU Symposuim on Computer Arabization, King Saud Univ*., Riyadh, Saudi Arabia (April 1987), 5-8.

[187] Tolba, M. F., Goneid, A., Salem, A. and Shaddad, E. "The Sensitivity of Some Recognition Algorithms for Printed Arabic Characters." *Proc. of the 13th IFIP Conference*, Tokyo, Japan, (1988), 735-743.

[188] Tolba, M. F. and Shaddad, E. "On the Segmentation and Recognition of Printed Arabic Characters." *Proc. of the Second Conf. on Arabic Computational Linguistics*, Kuwait, (Nov. 1989), 490-507.

[189] Tolba, M. F. and Shaddad, E. "On the Automatic Reading of Printed Arabic Characters." *Proc. of the IEEE Conference on Systems, Man and Cybernetics*, Los Angles: CA, USA, (1990), 496-498.

[190] Walker, R. F., Bennamoun, M., Boashash, B. "Comparative Results for Arabic Character Recognition Using Artificial Neural Networks." Presented at the *Workshop on Signal Processing and its Applications.*

[191] Yacu, S. G. *Design and Implementation of English Reading Machine for the Blind and Arabic Character Recognition*. M.Sc. Thesis, University of Baghdad, Iraq (1985).

[192] Yalabic, N. and Ozcilingir, M. "Computer Recognition of Ottoman Text." *Proc. of 3rd Int. Symp. on Computer and Information Sciences ISCIS III*, Cesme, Turkey,(Nov. 1988), 323-330.

[193] Yousef, M. I. *Recognition of Hand-Written Indian Numerals*. M.Sc. Thesis, Technology University, Baghdad, Iraq (in Arabic) (1986).

[194] Youssry, M. Youssef. "An Expert Knowledge-Based System for Arabic OCR." *Proc. of the 14th Conf. for Statistics, Computer Science, Social and Demographic Research*, Cairo, Egypt, (March1989), 227-245.

[195] Youssry, M. Youssef. "Arabic Optical Character Recognition (AOCR)." AlHandasah, Kuwait , No.12 (in Arabic) (Feb. 1993), 18-23.

[196] Zahour, A., Taconet, B. and Faure, A. "A New Method for Recognition of Arabic Cursive Scripts." *Proceedings of the IASTED International Symposium, Applied Informatics-AI '89*, Grindlewald, Switzerland, (in French) (February 1989), 97-101.

[197] Zahour, A., Taconet, B. and Faure, A. "A New Method for Recognition of Arabic Cursive Scripts." *Proc. First Int. Conference on Document Analysis & Pattern Recognition*, Saint-Malo: France (1991), 454-463.

[198] Zahour, A., Taconet, B. and Faure, A. "Machine Recognition of Arabic Cursive Writing ." *In: From Pixels to Features III: Frontiers in Handwriting Recognition*, Impedovo, S. and Simon , J. C. (Eds.), Amsterdam, Elsevier Science Publishers B.V. (1992), 289-296.

[199] Zaki, F. W., El-Konyaly, S., Abdel-Fattah, A. and Enab, Y. "A New Technique for Arabic Handwriting Recognition." *Proc. of the 11th Conf. for Statistics, Computer Science, Social and Demographic Research*, Cairo, Egypt, Vol. 2 (1986), 13.1.1-13.1.9.

التعرف على الحروف العربية : التقدم والتحديات

براويز أحمد و يوسف العوهلي

قسم علوم الحاسب، كلية علوم الحاسب والمعلومات، جامعة الملك سعود

ص .ب ٥١١٧٨ الرياض ١١٥٤٣ المملكة العربية السعودية

(قدّم للنشر في ١٩٩٨/٠٢/٢٤م؛ وقبل للنشر في ١٩٩٩/٠٣/١٦م)

ملخص البحث. يمكن لنظم التعرف على الحروف توفير الحلول لمشاكل إدخال البيانات إلى الحاسوب، التي تمثل عنق الزجاجة في كثير من مجالات التصنيع والمعالجة المعلوماتية. ولذلك؛ تمت دراسة وتطوير مثل هذه النظم لكافة اللغات العالمية، ويجب ألا تكون اللغة العربية مستثناة من ذلك. وخلال العقود الثلاثة الماضية، وجد عدد ملحوظ من الأبحاث وأعمال التطوير باتجاه الوصول إلى نظام فعال للتعرف على الحروف العربية. في هذه الورقة، نعرض مراجعة وتقويم شامل للطرق المستخدمة في تطوير أنظمة التعرف على الحروف العربية، إضافة إلى سرد مفصل لأهم المراجع في هذا المجال.