

Arabic Character Recognition using Approximate Stroke Sequence

Professor Mohammed Zeki Khedher*

Dr. Gheith Abandah¶

*Dept. of Electrical Engg khedher@ju.edu.jo

¶Dept. of Computer Engg abandah@ju.edu.jo

University of Jordan, Amman – Jordan

Abstract

Arabic character recognition of handwriting is addressed. A novel approach for the Arabic Character Recognition is presented based on statistical analysis of a typical Arabic text is presented. Results showed that the sub-word in Arabic language is the basic pictorial block rather than the word. The method of approximate stroke sequence is applied for the recognition of some Arabic characters in their stand-alone form. This method could be extended further for more accurate results. It is recommended that research in Arabic OCR systems in the future is based on the basis of the sub-word as the basic block rather than the word.

1. Introduction

Automatic recognition of handwriting has become a mature discipline at the beginning of the 21st century. On-line systems are now available on handheld computers with acceptable performance. Off-line systems are less accurate than on-line systems. However, they are now good enough for specialized systems such as interpreting handwritten postal addresses on envelopes and reading currency amounts on bank checks (Plamondon 2000).

The recognition of Arabic characters is particularly difficult due to the necessity of segmentation even for printed text. In order to get an insight into the Arabic word structure, it becomes necessary to do some statistical analysis on some typical Arabic text in order to assess the nature of problems facing the workers on Arabic OCR systems. For this purpose, a reasonable size of Arabic text was selected and analyzed. Based on the results of this analysis, a new procedure is suggested for building Arabic OCR systems. As a first step in the implementation of such systems, recognition of Arabic characters in their stand-alone form is addressed. The method of approximate stroke sequence matching is applied and the results are shown.

The paper gives some literature survey on previous work done in the field. It then gives the main characteristics of Arabic writing, presenting the importance of the sub-word structure of the Arabic word, showing the statistical results proving this phenomena, and proposing a new procedure for Arabic OCR system. The newly proposed method suggests the treatment of the sub-word as the basic block in the recognition of Arabic characters. The size of the sub-word should be treated as a decisive factor in the method of recognition of the characters contained in the sub-word. The method of approximate stroke sequence matching is described and then applied to an example of unknown character and compared with two standard characters. A text containing different shapes of Arabic characters was written by 48 different persons and samples of these characters under test were copied for this study. Some results of applying this procedure onto different characters is given. The paper discusses the results obtained and ends up with some conclusions and suggestions for future work.

2. Previous Work in Arabic Character Recognition

Several good literature review papers were published for various research topics on Arabic character recognition (Jambi, 1991; Ahmed, 1994 and Amin, 1997). A recent comprehensive one is given by Ahmed in 2000. Here we shall give some elaboration on some of the effort spent in this direction.

Classical optical off-line recognition of handwriting is composed of a pre-processing stage, character segmentation, feature extraction and a classification stage (Casey 1996). Pre-processing consists of several operations like thresholding, noise removal, page orientation, skewing of lines removal, line segmentation, word segmentation, and pictures and figures removal. There is little difference in these processes between Arabic language and Latin-alphabet-based languages (Abdulla, 1988; Mahmoud, 1991; Hussain & Cowel, 2000; Hussain & Zalik, 2000).

Work on isolated printed Arabic characters and numerals took a lot of shapes. Overall vertical level of the character compared with the baseline was studied (Talba 1987). The chain code describing the sequence of character strokes using the 8-direction strokes was followed by the majority of researchers (Alshebeli, 1997). However, hexagonal sampled procedure was also found (Khellah, 1994). Different methods for matching the unknown character with the standard characters were followed.

Segmentation of characters is an important step in character recognition for cursive writing whether hand written or printed. There are three strategies for segmentation: the classical approach in which segments identification is based on "character-like" properties, the recognition-based segmentation strategy, in which the system searches the image for components that match classes in its alphabets, and the third strategy is the holistic method, in which the system seeks to recognise words as a whole, thus avoiding the need to segment into characters (Casey 1996) Different algorithms were followed to apply one of the above methods or the other, like hierarchical syntactic procedure (Haj-Hassan, 1990), quadratic discriminating functions (Udpa, 1992), the method of moment invariant algorithm (El-Khaly, 1990), accumulative invariant moment was used as an identifier

in character recognition (El-Dabi, 1990), and even segmentation of printed Arabic characters was tried without the thinning process (El-Sheikh, 1988). Use of clustering technique was chosen for classification (El-Desouky, 1992) or tree representation for the description of various characters (Al-Waily, 1989; Saleh, 1994 & Saleh, 1996). Use of tree representation and fuzzy constrained graph models which tolerate large varieties in writing styles were reported also (Abuharba, 1994).

Recognition of different fonts of Arabic printed text was tried using pre-processing and structural feature extraction (Kavianifar, 1998). Parallel Arabic OCR systems were also proposed (Alherbish, 1997).

Hidden Markov models which proved to be very successful in the area of automatic speech recognition was tried in the area of omnifont, open-vocabulary Arabic OCR system (Bazzi, 1999).

Work on limited hand-written Arabic text database was tried. A system based on four types of basic features, namely the end points, corners, the strokes and the branch points gave reasonable results (Jambi, 1991).

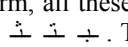
Recognition of Typewritten Arabic characters gave good results using external features such as character area ratio, n-th quadrant ratio, vertical line ratio, horizontal line ratio, number of upper edges, and other similar features (Al-Ohali, 1995).

Online character recognition uses the feature extraction process results in a sequential manner which is called the chain code. Treatment of secondary characters (mainly the points above and below the characters) is definitely an integral part of the recognition process (El-Gwad, 1990).

Neural Network was used in some work (Said, 1998, Al-Kadi, 1995, Altuwajri, 1995 and Al-Sharaidah, 2000).

3. Main characteristics of Arabic Writing

Arabic text is written from right to left and is always cursive. The shape of an Arabic character changes according to its location in the word. An Arabic character has up to four different shapes; the shape of a character depends on the type of character to its right and its position within the word. Table 1 shows the Arabic character set in the four different shapes.

The Arabic character set is composed of 28 basic characters. Fifteen of them have dots and 13 are without dots. Dots above and below the characters play a major role in distinguishing some characters that differ only by the number or location of dots. Take the example the letters: ب ت ث ي ن. In their middle form, all these five letters are written the same way as: . They differ only by the number or the locations of the dots.

There are four characters which may take the secondary character "Hamzah". Those are "Alif ا", "Waw و", "Yaa ي" and "Kaf ك".

There are also some other secondary characters used above and below the characters to indicate vowels but we shall exclude them now from our discussions.

Arabic characters do not have fixed width or fixed size, even in printed form.

3.1 An important phenomena in Arabic writing

Arabic writing is known to be cursive even in printed form. However, it differs from cursive handwriting of

Letter	Stand-alone	Initial	Middle	Final	Other shapes
Alef	ا			آ	ى
Ba'	ب	بـ	بـ	بـ	
Ta'	ت	تـ	تـ	تـ	ة
Tha'	ث	ثـ	ثـ	ثـ	
Jeem	ج	جـ	جـ	جـ	
H'a'	ح	حـ	حـ	حـ	
Kha'	خ	خـ	خـ	خـ	
Dal	د			ـد	
Thal	ذ			ـذ	
Ra'	ر			ـر	
Zai	ز			ـز	
Seen	س	سـ	سـ	سـ	
Sheen	ش	شـ	شـ	شـ	
Sad	ص	صـ	صـ	صـ	
Dhad	ض	ضـ	ضـ	ضـ	
Tta	ط	طـ	طـ	طـ	
Dha'	ظ	ظـ	ظـ	ظـ	
Ain	ع	عـ	عـ	عـ	
Gahin	غ	غـ	غـ	غـ	
Fa'	ف	فـ	فـ	فـ	
Qaf	ق	قـ	قـ	قـ	
Kaf	ك	كـ	كـ	كـ	
Lam	ل	لـ	لـ	لـ	
Meem	م	مـ	مـ	مـ	
Noon	ن	نـ	نـ	نـ	
Ha'	هـ	هـ	هـ	هـ	
Waw	و			ـو	
Ya'	ي	يـ	يـ	يـ	

Table1: The Different Forms of Arabic Alphabets

English in that some characters can be connected from one side only. Out of the 28 basic Arabic characters, six can be

connected from the right side only while the other 22 can be connected from both sides. These six characters are: dal (د), raa (ر), waw (و), alef (ا), thal (ذ), and zay (ز). These six characters have only two forms, the stand-alone form and the final form. Whereas the rest of the characters can appear in any of four forms: the initial, the middle, the final, and the stand-alone form. Consequently, an Arabic word may consist of one or more sub-words. A sub-word can be defined as the basic stand-alone pictorial block of the Arabic writing. Any optical character recognition of Arabic characters should treat the sub-word as the basic block for processing whatever the method it uses for preprocessing, segmentation, recognition, or classification.

This is because each sub-word is separated from other sub-word by a space. Although spaces between sub-words are usually shorter than those between successive words, still they are surrounded by space. A word may contain one or more sub-words. Some of these sub-words may even consist of a single character in its stand-alone form. Hence, their recognition does not need segmentation.

Shape of the letter in the text differs according to the location of the character in the sub-word, i.e. a character at the end of sub-word, has exactly the same shape when it comes at the end of a full word. Take the example of the

Char per sub-word	Sub-words	Sub-words %	Stand-alone characters	Initial characters	Middle characters	Final characters
1	263,065	45.80%	263,065	0	0	0
2	159,995	27.90%	0	159,995	0	159,995
3	90,068	15.70%	0	90,068	90,068	90,068
4	43,124	7.50%	0	43,124	86,248	43,124
5	13,433	2.30%	0	13,433	40,299	13,433
6	3,633	0.63%	0	3,633	14,532	3,633
7	818	0.14%	0	818	4,090	818
8	247	0.04%	0	247	1,482	247
Total	574,383	100.00%	263,065	313,318	236,719	313,318
% characters			23.4%	27.8%	21%	27.8%

Table 2: Sub-words and Shapes Statistics

word رجال . It consists of 3 sub-words, containing 1, 2, and 1 character respectively.

3.2 Test Sample and Results

In order to give a fair idea about sub-words, a sample of Arabic text consisting of about 1.4MB was collected. It was randomly selected from old books, modern books, newspapers, and other available sources on the web. Statistics presented here about this sample text may give an idea about the structure of Arabic words in terms of sub-words and the four character shapes. Table 2 shows the analysis for this sample.

The sample consists of 262,647 words with 1,126,420 characters. This means that the average word length is 4.3 characters per word. The number of sub-words is 574,383. This means that on the average there is 2.2 sub-word per word. The number of sub-words consisting of one character is 263,065 which makes about 45.8% of the total number of sub-words (and 23.4% of the total number of characters). This means that, in the process of optical character recognition, slightly less than one half of the sub-words need no segmentation at all (whether printed or handwritten). The number of sub-words that consist of two characters is 159,995 which makes 27.9% of the total number of sub-words. This means that about 30% of the total number of sub-words need segmentation into two characters only. The table also shows that on the average the four different shapes of characters are almost equal with the middle form slightly less (23.4% stand-alone, 27.8% for each of the initial form and the final form, as the number should be equal, and 21% for the middle form).

3.3 Proposal for a New Procedure for Recognition of Arabic Characters

According to the above discussion, the approach suggested here is to separate the text into three groups:

1. Sub-words consisting of one character that is in the stand-alone form. This is to be recognised directly without any segmentation.

2. Sub-words consisting of two characters. The first one is in the initial form and the second one is in the final form. This needs segmentation in two parts only. If there is a pre-knowledge of the number of characters to be segmented in the sub-word, then the task becomes easier.

3. Sub-words consisting of more than two characters. The first one is in the initial form, the last one in the final form, and the rest are in the middle form.

Figure 1 shows the flow diagram for this procedure.

4. Approximate Stroke Sequence String Matching

Given two character images, there is no universally accepted definition for similarities or differences between them. If the two images are converted into a one-dimensional string, then the task will be easier to define. Distance between two histograms of angular measurements was the subject of some literatures (Cha 2000).

The stroke sequence is based on the 8-direction stroke convention shown in Figure 2.

Stroke sequence string matching (Cha 1999) is based on the individual distance $d_{i,j}$ between the i 'th stroke in letter a_1 and the j 'th stroke in letter a_2 where

$$d_{i,j} = |a_1(i) - a_2(j)| \quad \text{if } |a_1(i) - a_2(j)| \leq 4$$

This value is to be modified so that the $d_{i,j}$ value is replaced by the value $8-d_{i,j}$ when it exceeds the value of 4. This is equivalent to taking the smallest angle between the two directions whether the rotation is clockwise or counter-clockwise. Hence $d_{i,j}$ gives the minimum number of necessary steps to turn from the direction given by $a_1(i)$ and $a_2(j)$.

Allowing a cost function of $c=2$ which allows a mini-

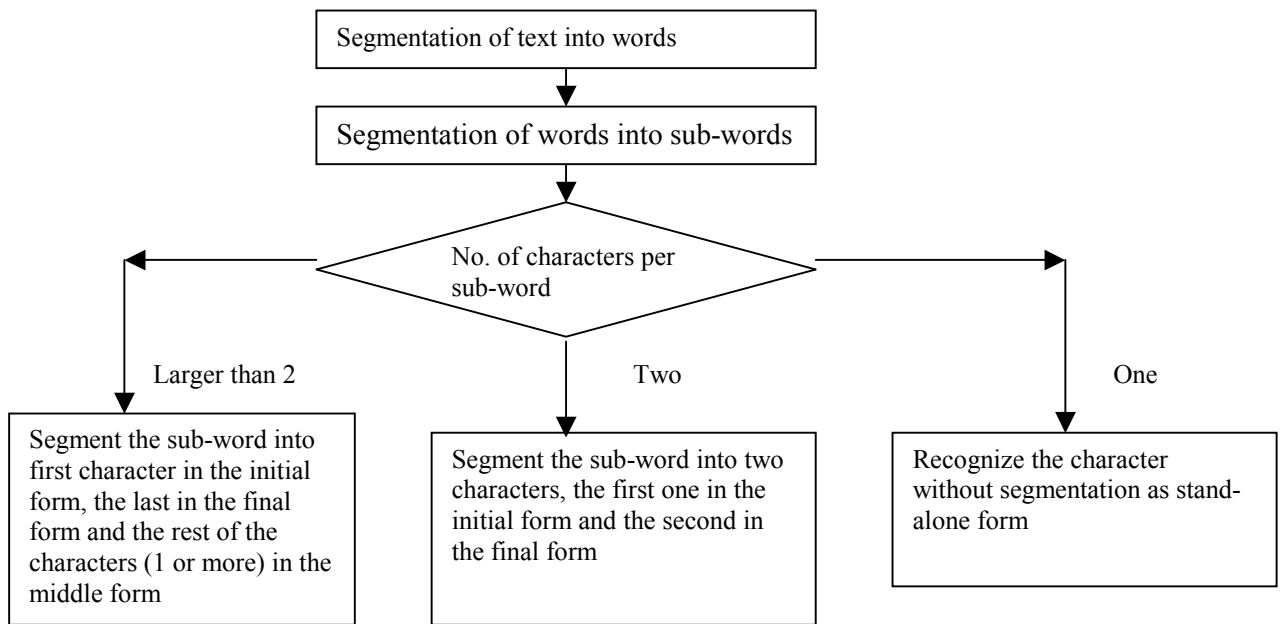


Figure 1: Proposed Procedure for Arabic Character Recognition

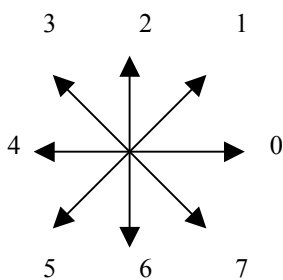


Figure 2: The 8-direction stroke convention

minimum edit distance between two stroke sequence strings.

Figures 3a and 3b show the Arabic characters ζ and ξ taken as standard shapes. The stroke sequences corresponding to each of them is shown in the upper

row of Tables 3 and 4, respectively. Figure 3c shows a supposed unknown character to be compared to each of them. However, this is a hand-written letter ξ . The stroke sequence for this character is shown in the left-hand column of both Tables 3 and 4. They also show the computed distance tables between the unknown character and each of the two characters which are nearest in shape to it. Its distance to the letter ξ is 8 while its distance to the letter ζ is 10. This gives the result that the unknown character is ξ and not ζ .

The individual strokes on the top and left hand side of the table (starting row and column) are t_i and l_j simultaneously. The calculations involved in the shown Tables takes into consideration that

$T(i,j)$ is the minimum value of $T(i-1, j-1) + d_{i,j}$ and the two values $T(i-1, j) + c$ when t_i is missing and $T(i, j-1) + c$ when l_j is missing.

		2	0	0	5	5	6	7	0	1	2
	0	2	4	6	8	10	12	14	16	18	20
2	2	0	2	4	6	8	10	12	14	16	18
4	4	2	4	6	5	7	9	11	13	15	17
5	6	4	5	7	6	5	7	9	11	13	15
6	8	6	6	7	8	7	5	7	9	11	13
0	10	8	6	6	8	9	7	6	7	9	11
4	12	10	8	8	7	9	9	8	9	10	11
5	14	12	10	10	8	7	9	10	11	12	13
6	16	14	12	12	10	9	7	9	11	13	15
7	18	16	14	14	12	11	9	7	10	12	14
0	20	18	16	14	14	13	11	9	7	9	11
2	22	20	18	16	16	15	13	11	9	8	9
4	24	22	20	18	17	17	15	13	11	10	10

Table 3: String matching between the unknown character and character ζ

		3	4	5	6	7	0	4	5	5	6	7	0	1
	0	2	4	6	8	10	12	14	16	18	20	22	24	26
2	2	1	3	5	7	9	11	13	15	17	19	21	23	25
4	4	3	1	3	5	8	9	11	13	15	17	19	21	23
5	6	5	3	1	3	5	7	9	11	13	15	17	19	24
6	8	7	7	3	1	3	5	7	9	11	13	15	7	19
0	10	9	9	5	3	2	3	5	7	9	11	13	15	17
4	12	11	9	7	5	4	5	3	5	7	9	11	13	15
5	14	13	11	9	7	6	7	5	3	5	7	9	11	13
6	16	15	13	11	9	8	8	7	5	4	5	7	9	11
7	18	17	15	13	11	9	9	9	7	6	5	5	7	9
0	20	19	17	15	13	12	11	11	9	8	7	6	4	7
2	22	21	19	17	15	14	13	13	11	10	9	8	7	6
4	24	23	21	19	17	16	15	13	13	12	11	10	9	8

Table 4:String matching between the unknown character and character

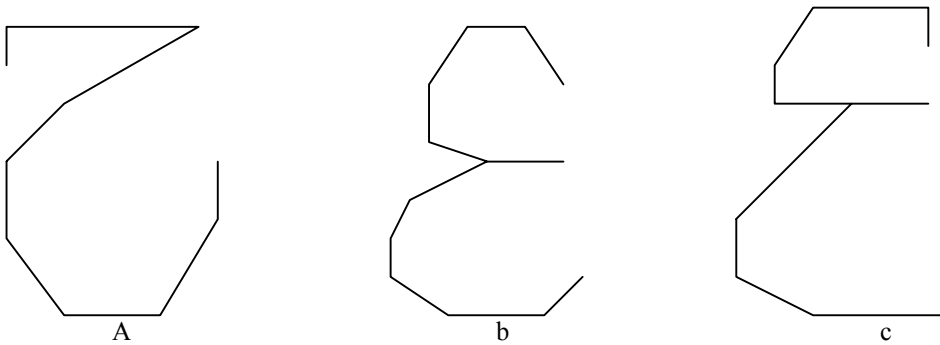


Figure 3: a: Character ح b: character ع d: unknown character to be matched with a and b



Figure 4: The letter ع hand written by 48 different persons

1. Results and Discussion

The above algorithm was applied to various Arabic stand-alone characters collected from 48 different persons. Figure 4 shows such character from the

different handwriting of the 48 persons for the letter ع. These persons were asked to copy the same text without any restrictions on their writing style. The handwritten pages were scanned and then normalized to nearly the same overall sizes. The stand-alone characters were then copied and analysed by the above algorithms. Results showed recognition rate of about 80% for characters such as ا, ص, ه while the rate of recognition was much less than that for similar characters such as ب & ف or ق & ن.

6. Conclusions

A statistical analysis on a sample of Arabic text, showed that the average Arabic word contains about 4.3 characters with an average of 2.2 sub-words. This shows that the basic block to be dealt with in Arabic OCR systems should be the sub-word rather than the word. The size of the sub-word varies from a single character up to 8 characters. The method of recognition of sub-words of different lengths ought to be different. For sub-words with a single character, the matching for recognition may be made only with characters of the stand-alone form. For sub-words with two characters only, a single shot segmentation has to be made dividing the sub-word into two characters. The first one is in the initial form and the second one in the final form. Sub-words of lengths longer than 2 characters need to be segmented into three characters or more. The first is of initial form, the last of final form and the rest

of middle form. Design of Arabic OCR system when taking these facts into account would be much simpler. However, the classification of sub-words according to the number of the characters they contain, still ought to be addressed.

As a first step, the recognition of the sub-words with a single character of a stand-alone form has been treated using the approximate stroke sequence string matching. Promising results are shown. Further refinement of the algorithm used need to be carried out for better rate of recognition.

7. Acknowledgement

The authors are grateful to the grant offered by The Higher Council for Science and Technology, Jordan and to the University of Jordan for offering the facilities to do the research. The authors also thank Mr. Samer Alayysh for helping in getting some of the results

8. References

- Abd El-Gawad A.O., M.M. Salem, F.E.Z. Abou Shadi and H. Arafat, 1990. Automatic Recognition of Handwritten Arabic Characters. *10th International Conference on Pattern Recognition*.
- Abdulla W.H., A.O.M. Saleh and A.H.Morad, 1988. A Preprocessing Algorithm for Hand-Written Character Recognition. *Pattern Recognition Letters*. 7, 13-18.
- Abo Samra G., K. Jambi, H. Al-Barhamtoshy, R. Amer and I. Al-Bidewi, 1997. A Comprehensive Algorithm for Segmenting Handwritten Arabic Script in Off-line Systems. *17th NCC, KFUPM*. 1-13.
- Abuharba I.S.I., S.A. Mahmoud and R.J. Green, 1994. Recognition of Handwritten Cursive Arabic Characters. *HLL Transaction of Pattern Analysis and Machine Intelligence*.16:6. 664-672.
- Ahmed P. and Y. Al-Ohali, 2000. Arabic Character Recognition: Progress and Challenges. *Journal of King Saud University*. 12, Comp. & Inf. Sci. 85-116
- Ahmed P. and M. A. A. Khan, 1994. Computer Recognition of Arabic Script Based Text: the State of the Art. *Proc. of the 4th International Conference and Exhibition on Multi-lingual Computing (ICEMCO 94, Arabic and Roman Script)*, London, 22.1-22.15..
- Al-Ohali Y.A.M, 1995. *Development and Evaluation Environment for Typewritten Arabic Character Recognition*. M.Sc. Thesis, King Saud University. Riyadh.
- Al-Sharaidah, M. A., 2000. *Recognition of Handwritten Arabic Characters via Neural Networks* M.Sc. Thesis, Jordan University.
- Al-Yousefi H. and S.S. Udpa, 1992. Recognition of Arabic Characters," *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 14: 8. 853-857.
- Altuwijiri M.M. and M.A. Bayoumi, 1995. A New Recognition System for Multi-Font Arabic Cursive words. *Proceedings of ICECS'95*, Amman, Jordan. 298-303.
- Alshebeili S.A., A.A.F Nabawi and S.A. Mahmoud, 1997. Arabic Character Recognition Using 1-D Slices of the Character Spectrum. *Signal Processing, Elsevier*. 56. 59-75.
- Alherbish J. and R.A. Ammar, 1997. Arabic Character Recognition in a Multi-processing Environment. *Proceedings of the 2nd IEEE Symposium on Computers and Communications*.
- Al-Kadi Z. and S. Serhan, 1994. The Use of Learning Networks in Recognition of Arabic Numeral. *Dirasat*, Jordan University. 22B:4. 933-949.
- Al-Waily R.S.A., 1989. *A Study on Preprocessing and Syntactic Recognition of Hand-Written Arabic Characters*. M.Sc. Thesis, Basrah University, Iraq.
- Amin A., 1997. *Handbook of Character Recognition and Document Image Analysis*, Chapter 15 World Scientific Pub. Co. 397-420.
- Bazzi I., R. Schwartz and J. Makhoul, 1999. An Ominfont Open-Vocabulary OCR System for English and Arabic. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 21: 6. 495-504.
- Casey R.G. and E. Lecolinet, 1996. A Survey of Methods and Strategies in Character Segmentation. *IEEE Transaction on Pattern Analysis And Machine Intelligence*. 18:7.
- Cha S.H., Y.C. Shin and S.N. Srihari, 1999. Approximate Stroke Sequence String Matching Algorithm for Character Recognition and Analysis. *5th International Conference on Document Analysis & Recognition*. Bangalore, India
- Cha S.H. and S.N. Srihari, 2000 "Distance between Histograms of Angular Measurements and its Application to Handwritten Character Similarity," *Proc. of International Conference on Pattern Recognition*.
- El-Desouky A.I., M.M. Salem A.O. Abd El-Gwad and H.Arafat, 1992. A Handwritten Arabic Character Recognition Technique for Machine Reader. *International Journal of Mini and Microcomputers*. 14:2. 57-61.
- El-Khaly F. and M.A. Sid-Ahmed, 1990. Machine Recognition of Optically Captured Machine Printed Arabic Text. *Pattern Recognition*. 23: 11, 1207-1214.
- El-Sheikh T.S. and R.M. Guindi, 1988. Computer Recognition of Arabic Cursive Script. *Pattern Recognition*. 21:4. 293-302.
- Hussain F., B. Zalik and S. Kolmanic, 2000. Intelligent Digitisation of Arabic Characters. *Proceedings of the Conference on Information Visualization*. 337-342.
- Hussain F. and J. Cowell, 2000. Character Recognition of Arabic and Latin Scripts. *Proceedings of the Conference on Information Visualization*. 51-56.
- Haj-Hassan F., 1990. Arabic Character Recognition. *Computers and the Arabic Language*, Hemisphere Publishing Co. 113-118.
- Jambi K. M., 1991. Arabic Character Recognition: Many Approaches and One Decade. *The Arabian Journal for Science and Engineering*. 16: 4B. 499-508.
- Jambi K.M., 1991. *Design and Implementation of a System for Recognizing Arabic Handwritten Words with Learning Ability*. Ph.D. Thesis, Illinois Institute of Technology.
- Kavianifar M. and A. Amin, 1998. Preprocessing and Structural Feature Extraction for a Multi-Font Arabic/Persian OCR. *Proceedings of 5th International Conference of Document Analysis and Recognition*.
- Khellah F., 1994. Recognition of Hexagonally Sampled Printed Arabic Characters. *The Arabian Journal for Science and Engineering*. 19:4A. 565-585.

- Mahmoud S. A., I. AbuHaiba and R.J. Green, 1991 Skeltonization of Arabic Characters Using Clustering Based Skeltonization Algorithm. *Pattern Recognition*. 24: 5. 453-464.
- Nouh A., A.N. Ula and A. Sharaf Eldin, 1988 Algorithm for Feature Extraction: A case Study for the Arabic Character Recognition. *Proceedings of 10th National Computer Conference*, Saudi Arabia. 653-666.
- Plamondon, R., 2000. On-Line and Off-line Handwriting Recognition, A Comprehensive Survey. *IEEE Transaction on Pattern Analysis And Machine Intelligence*, 22:1.
- Said F. N., R.A. Yacoub and C.Y. Suen, 1998. Recognition of English and Arabic Numerals Using a Dynamic Number of Hidden Neurons. *Proceedings of 5th International Conference of Document Analysis and Recognition*.
- Salih A.O.M. and R.S. Al-Waily, 1994. Tree Representation of Hand-written Arabic Characters. *Mu'tah Journal for Research and Studies*. 9:3. 125-140.
- Salih A.O.M. and R.S. Al-Waily. 1996. Recognition of Hand-Written Arabic Characters sing Tree Grammars 2nd *Jordanian Conference in Electrical Engineering, Mutah University*. 402-413.
- Sami S.E., R. Ramsis and A., 1990. Arabic Character Recognition System, A Statistical Approach for Recognizing Cursive Typewritten Text. *Pattern Recognition*. 23: 5, 485-495.
- Talba M.F., S.A. Wahba and A. Salim, 1987. A Recognition Algorithm for Arabic Printed Characters," *Proceedings of the International Symposium of Applied Informatics*, Switzerland. 128-131