Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 5, May 2017, pg.4 – 14

Review about Off-line Handwriting Arabic Text Recognition

Wafa Alameen Alsanousi[1], Iman Siddig Adam[1], Mohsen Rashwan[2], Shrief Abdou[3]

[1]College of Computer Science and Information Technology, Sudan University of Science and Technology, Sudan, wafa.alameen@hotmail.com, imanishappy@gmail.com

[2]Prof. Dr. Faculty of Engineering, College of Electronics and Electrical Communications, University of Cairo, Giza, Egypt, mrashwan@rdi-eg.com

[3]Dr.Faculty of Engineering, College of Electronics and Electrical Communications**,** University of Cairo, Giza, Egypt, sh.ma.abdou@gmail.com

Abstract: This study covers the Optical Character Recognition (OCR) systems for off-line type. This paper focus on the Handwriting Arabic Text Recognition because technology is still not mature enough to deal with Arabic handwriting, compared with the writings of the Latin alphabet. In this study, different approaches used for off-line handwritten Arabic word recognition were reviews. The main stages of handwritten Arabic OCR systems and the major technical challenges of these systems are presents, provides a comprehensive review of the research done in this field. Also we focus on initial stages, including the pre-processing and segmentation stages since they are important stages in building of OCR system. A review of previous work in OCR systems is conducted. Also, we presented a database summary of Arabic handwritten texts and images.

Keywords: offline Arabic Handwriting Recognition, OCR,, Pre-processing, Segmentation, PAWs.

1. Introduction

The Optical Character Recognition (OCR) is an important filed in pattern recognition for images. OCR is the process of detecting and recognizing characters from input image and converting it into ASCII or other equivalent editable machine form [**1**].

OCR systems initially focused on printed text due to its ease of recognition, but over the last few decades Handwritten Character Recognition (HCR) has become feasible and very important for the creation of electronic libraries, digital copies of handwritten documents, writer identification, automatic mail sorting, cheques' processing, and data entries [**2**].
The Arabic language is one of the popular script languages in the world. It is a semantic language with a composite morphology. Working on Arabic OCR started in the 1970s [**3**]. Arabic text is considered as a rich language and contains huge number of words [**4**]. However, there has been a lack of effort in the recognition of Arabic OCR and handwritten system compared to the recognition typewritten documents.

In this paper, different approaches used for handwritten Arabic word recognition were reviewed. Also, this paper reviewed the main stages of handwritten Arabic OCR systems, presented the major technical challenges of these systems, and provided a comprehensive review of the research done in this field.

The organization of the paper is as follows: section 2 gives an overview of the characteristic of Arabic language, section 3 defines and introduces the handwritten recognition; section 4- demonstrates the stages of the optical character recognition system including the preprocessing. That consists of: segmentation, binarization, noise removal, page orientation, pictures and figures removal, baseline detection, skew/slant detection and correction. Section 5- is focusing on the databases for Arabic handwriting words and images and the last section, section 6-contains the conclusion.

2. Characteristic of Arabic script

Arabic language is widely spoken by more than 23 countries in Middle East and North Africa, written by more than 340 million people [**5**]. Arabic language has challenging characteristics that differ from the Latin languages. These challenges make the recognition rate much lower compared to the Latin language. Many languages such as Persian, Urdu and Jawi, use Arabic characters. Nevertheless, compared to Latin and Chinese handwritten character recognition, little research has been done into Arabic handwritten character recognition [**6**] [**7**].

Arabic text use diacritics marks that help in resolving linguistic ambiguity of the text [**4**] [**8**].

Arabic language is written in horizontal lines from right to left, it contains 28 characters and it is always cursive in both machine printed and handwritten form. That means the segmentation of printed Arabic character is not easy compared with Latin printed text [**9][10**]. Sixteen characters of Arabic contain dots, which are one, two or three, under or above the characters. Arabic letters are written in either continuous or isolation form. Most of the Arabic characters vary in shape depending on their position in the word (at beginning, in the middle, at the end or in isolation) [**11**], (**See Table. 1).**

For OCR systems the Arabic language has several challenges such as the multiple grapheme shapes, the ligatures, the characters overlapping, the diacritics marks, and the dots. Figures (1) and (2) show sample Arabic script with example cases for OCR system challenges.

3. Handwritten Recognition

Handwriting text recognition systems differ in getting entered. There are two types, either off-line (Machine printed and handwritten) or on-line (Isolated character and cursive words). **[12**]

Off-line character recognition takes a raster image from a scanner (scanned images of the paper documents), digital camera or other digital input sources. The image is binarized through threshold technique based on the color pattern (color or gray scale), so that the image pixels are either 1 or 0. The off-line recognition of texts can be further divided into two categories: the recognition of printed and of handwritten characters. Printed characters have one style and size for any given font. However, handwritten characters have styles and sizes which vary, both for the same writer and between different writers. Handwritten words can be recognised in two ways: recognition of a whole word without segmentation, or recognition based on segmentation [**7**].

In on-line, the current information is presented to the system and recognition (of character or word) is carried out at the same time. Basically, it accepts a string of (x, y) coordinate pairs from an electronic pen touching a pressure sensitive digital tablet [**13**].

Despite 35 years of research and development to build OCR systems, such technology is still not mature enough to deal with Arabic handwriting, compared with the writings of the Latin alphabet. However, there are still improvements in terms of reducing the error rate of the words, the ability to work in the presence of a reasonable amount of noise and dealing with various fonts and sizes, as well as non-compliance with certain linguistic results [**14**].

Table (1): Arabic Characters shapes

No	Name	Isolated	Connected		
			Beginning	Middle	End
1	ALIF	أ	أ	ـأ	ـا
2	BAA	ب	بـ	ـبـ	ـب
3	TAA	ت	تـ	ـتـ	ـت
4	THAA	ث	ثـ	ـثـ	ـث
5	JEEM	ج	جـ	ـجـ	ـج
6	HAA	ح	حـ	ـحـ	ـح
7	KHAA	خ	خـ	ـخـ	ـخ
8	DAL	د	د	ـد	ـد
9	THAL	ذ	ذ	ـذ	ـذ
10	RAA	ر	ر	ـر	ـر
11	ZAAY	ز	ز	ـز	ـز
12	SEEN	س	سـ	ـسـ	ـس
13	SHEEN	ش	شـ	ـشـ	ـش
14	SAAD	ص	صـ	ـصـ	ـص
15	DADD	ض	ضـ	ـضـ	ـض
16	TAH	ط	طـ	ـطـ	ـط
17	THAH	ظ	ظـ	ـظـ	ـظ
18	AIN	ع	عـ	ـعـ	ـع
19	GHEN	غ	غـ	ـغـ	ـغ
20	FAA	ف	فـ	ـفـ	ـف
21	QAF	ق	قـ	ـقـ	ـق
22	KAF	ك	كـ	ـكـ	ـك
23	LAM	ل	لـ	ـلـ	ـل
24	MEEM	م	مـ	ـمـ	ـم
25	NOON	ن	نـ	ـنـ	ـن
26	HAA	ه	هـ	ـهـ	ـه
27	WAW	و	و	ـو	ـو
28	YAA	ي	يـ	ـيـ	ـي

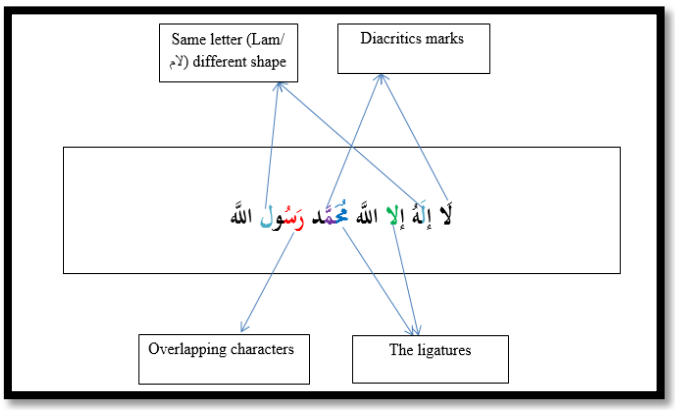

Figure (1): Some Arabic Characters challenges

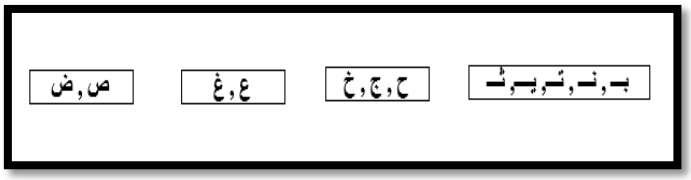

Figure (2): Some sets of graphemes are differentiated by dots

4. Optical Character Recognition

The main objective of OCR systems is to convert the image into machine-readable text. [**12**].

Generally, OCR system consists of five stages namely: pre-processing, segmentation, feature extraction, classification and post-processing as shown in **Figure (3)**. However we focusing on the initial stages; the pre-processing and the segmentation stages because they are important stages in building of OCR system.

The initial step in OCR system is capturing the input text data and converting it into a digital readable form [**12**].The Off-line character system recognizes the text after it has been written and is inserted by a video camera or scanner, the scanner is very suitable, because it has less noise at the imaging process and has other features such as automatic binarization **[9][15].**

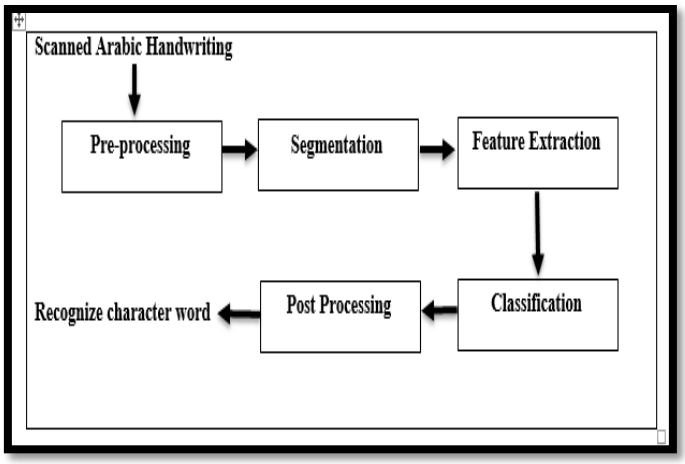

Figure 3: OCR System Architecture

4.1 Preprocessing

Preprocessing is the first step in OCR which consists of several operations performed to improve the quality of original image and make it suitable for future processing [**16**].

These operations usually include binarization, noise removal, page orientation, pictures and figures removal, baseline detection, skew/slant detection and correction, and resizing [**2][17] [18**]. There is little difference in these processes between Arabic language and other languages. Also in Arabic language there is a difference between off-line and off-line system

4.1.1 Binarization

The binarization or thresholding is the conversion of a color or grayscale image into binary image. Binary morphological operations like opening, closing, thinning and whole filling could be applied to enhance visibility and structural information of character [**19**]. Otsu's [**20**], proposed a global binarization technique, which chooses the threshold to minimize the interclass variance of the black and white pixels.

4.1.2 Noise removal

This processing operation identifies and removes pixels, which represent noise that do not belong to the word shape and may have been generated during document generation. Proper filter like mean filter, min-max filter, Gaussian filter,.. etc, could be applied to remove noise from documents. Those techniques can be categorized in two major groups; filtering and smoothing. Zhixin et al. [**21**] proposed four Algorithms for noise (Clutter noise, paper noise, form rule-lines noise, salt noise and broken stroke noise). Moreover, Ahmed et al. [**22**] discussed two types of technique used for noise removal.

I. ***Statistical noise removal*:** These techniques improve the image using contrast and dynamic range modification by reducing degradations that may be present, while median filtering is a nonlinear process that could be useful in preserving edges in an image by reducing random noise.
II. ***Morphological noise removal*:** This includes *Filling*; which fills isolated interior pixels, *Bridging*: that bridges unconnected pixels, *Removing*; that removes isolated pixel, and *Dilation*: an operation that grows or thickens objects in a binary image.

4.1.3 Baseline detection

The baseline is a virtual line on which the characters of Arabic text are aligned. Some techniques of finding the baseline include:

I. ***Horizontal histogram***: Srihari et al.[**23**] that finds the page skew by taking horizontal histogram across the image, and calculating the number of foreground pixels.
II. ***Skeleton-based Methods***: Azizi et al.[**24**] that produces a polygonal approximation to the skeleton of the word, while Boukerma et al. [**25**] Developed baseline estimation algorithm based on Sub-Words as a part of Arabic words (PAWs). However, Parvez et al. [**26**] has used more complex methods to adapt the expectation-maximization (EM) algorithm for baseline detection.

4.1.4 Slant angle correction

If a document is scanned, then it may not be perfectly horizontally aligned, Accorchngly, we need to align it by performing slant angle correction.

Parvez et al. [**26**] have proposed a method to calculate Near-Vertical Strokes (NVSs) by multi-direction Prewitt filters to estimate the slant angle. Ahmed et al. [**22**] have also proposed another approach. It is basic idea is to locate near-vertical strokes in the character and estimate the average slant of the character from these strokes. Thus, the slant in a character is corrected by applying a shear transformation to that character.

Steinherz et al. [**27]** have proposed a method based on principal component analysis (PCA) which is a way of finding the directions that the distribution exhibits the greatest variation among them. In order to overcome the computational effort of PCA, Azizi et al. [**21**] proposed an algorithm for skew correction based on finding the peaks and valleys to estimate the skew angle.

4.2 Segmentation

Segmentation is an important step in OCR system for cursive writing whether handwritten or printed. It is the separation of an image into basic units of different types that can simply be pixel segments using over-segmentation techniques, piece of Arabic words (PAWs) (vertical projection) or line segmentation (Horizontal projection). There are three strategies for segmentation: (1) the classical strategy, in which, segments identification is based on "character-like" properties. (2) The recognition-based segmentation strategy, in which the system searches the image for components that match classes in its alphabets, and (3) the holistic strategy, in which the system seeks to recognize words as a whole, thus avoiding the need to segment into characters. [**28][29][30**]. Many algorithms designed for Latin cursive word segmentation can be used for Arabic handwriting However, they are not adequate for that task, and for this reason, researchers proposed several modifications for the segmentation algorithms to be more suitable for Arabic handwriting **[31] [32].**

The most obvious efforts can be briefed as follows:

Jafaar and Xue**[33],** proposed new segmentation method which is entirely different from the existing retail methods. The results in terms of quality and performance were superior to retail methods used. presented method has experienced three major phases, firstly make the picture a thin of representation skeletonized by one pixel, secondly is transferred each pixel is diagonally to the nearest horizontal or vertical line, finally orthogonal encoding, these lines are a unique kind of valid numbers, each vector

representing one character of a single word. The system has been tested on the IFN/ENIT database. The accuracy is higher than (93.30%) compared with other system methods (89.10% to 90.26%). However the rules are hard to be understood by human.

Jawad et al, [**6**] used baseline detection and optimal thresholding for words segmentation in pre-processing stage of handwritten Arabic text. The system has been tested on the IFN/ENIT database. However, Word segmentation resulted good for two words but it was bad for one word and more than two words that mean the baseline detection and optimal thresholding techniques are very good in segmentation words for two words only.

Safwanet et al. [**34**] proposed algorithm for segmentation of Handwritten Arabic words based on both Contour and Skeleton Segmentation. The algorithm segments connected letters into smaller letters so that none of them should exceed three letters. The algorithm has been applied over 6300 words from 45 different documents from DARPA and IFTN databases. 90.4% of segments included one letter and 9.6% of the segments contain more than one letter.

Manal et al. [**35**] used word segmentation to recognize off-line Arabic handwriting character. They presented a simple methodology to authenticate Arabic handwriting character. The system is built with two stages pre-processing and then segmentation in first stage and recognition in the second stage. They concluded that the use of a fixed width gives the low recognition rate. But when they changed the width of the piece of the character, a better recognition rate 81% was noticed. The database consisted of Arabic letters written by two writers with different position shapes.

4.3 Feature extraction

The target of features extraction operation is to extract from the raw data the most suitable information for the classification purposes [**36**]. The selection of good features must take consideration the following criteria:

(1) Features prefer to be independent in size and rotation.

(2) Features must be computed easily and should be chosen, so they do not repeat each other **[12].**

4.4 Classification

Classification in the OCR system is the stage of making key decisions. Depending on the features extracted from the previous steps, the seed pattern that represents the input features could be determined [**12**]. The four main techniques used in classification stage are:

- Artificial Neural Network [**37**] [**38**], [**39**].
- Hidden Markov Model [**40**] [**41**], [**42**].
- K-nearest neighbour [**43**].

Khaoula et al. [**44**] used neural networks with feed-forward back propagation algorithm .Finally, the recognition of the character through the use of Neural Network techniques. The proposed approach is tested on 1400 different characters written by ten users. Each user wrote 28 Arabic characters five times in order to get different writing variations. The results obtained good result with the average accuracy of 83%.

Ahlam et al. [**45**] used multi stream approach (multi stream is combine the outputs of classifier) to recognize the offline handwritten Arabic word by using hidden Markov models (HMM), the steps are pre-processing, segmentation stage divided the text into lines and lines into words, followed by the most important stage is extracting features stage, that used of two methods to extract features sliding windows, horizontal and vertical and diagonal. The last stage is the recognition word stage using multi stream approach by used the hidden Markov models (HMM). The results obtained when using first method is equal 78.2% and second method equal 76.6% for recognition rate, but when they combine the two methods the rate increased to 83.8%.

Jawad et al. [**46**] used to classify the words using the K nearest Neighbour classifier (KNN). The proposed system has been successfully tested on the IFN/ENIT database consisting of 32492 Arabic

handwritten words which are written by more than 1000 different writers. Experimental results show a good recognition rate when compared with other methods.

4.5 Post–processing

The Post–processing stage is used to reduce word recognition error rate by correcting the output of the OCR system on the basis of a good working knowledge of the language [**12**]. The linguistic level could be either on character level, lexical level, morphological level, the syntactic, higher semantic level or discourse levels [**43**].

5. Databases for Arabic Handwritten words

Most of researchers used different printed Arabic texts databases. There are several Arabic text and words databases to serve handwritten Arabic characters, digit, word and text recognition research. In developing an OCR system for Arabic it is necessary to create a database of Arabic words. Such a database has many uses as well as in training and testing a recognition system.

The CENP ARMI database was developed by the centre for pattern recognition and machine intelligence. This database was used by researchers in field of Arabic handwritten recognition and consists of 2499 words, 29498 samples of Arabic sub- words and 15175 Indian digits. The database is divided randomly into either training or test sets. The training set includes 66-75% of the available data [**47**].

The IFN/ENIT database is the common database used by researchers who are work in field of Arabic handwriting recognition. This database was collected by the Institute of Communication Technology (IFN) and Ecole Nationed' Inge'nieurs de Tunis (ENIT). IFN/ENIT database was collected from 411 writers and consists of 26459 handwritten Tunisian towns, 115585 pieces of Arabic words (PAWs) and 212211 characters [**48**]. On the other hand in the past few years, several competitions have been managed by using this database [**49][50][51][52][53]**, **[54].**

Al-ISRA database was collected in Al-Isra University in Jordan. It contains 37,000 Arabic words, 10,000 digits, 2,500 signatures, and 500 free-forms of Arabic sentences all gathered from five hundred students [**55**].

Another database the AHDB database has contained the Arabic text and words that have been collected from more than 100 different writers. This database has included the most commonly used Arabic words, which are used in the writing of cheques's and handwritten pages [**56**].

The Arabic Printed Text Images (APTI) database was generated using 113,284 words, 10 font sizes, 10 Arabic fonts and 4 font styles. The database included 45,313,600 word images to spoken to more than 250 million characters [**57**].

Another database collected by SAIC [**58**], the DAMP database DARPA Arabic Machine Print document was used 297 images scanned from books, magazines and newspapers. This database was divided into three sets, 60 images for development, 60 images for testing, and 177 images for training.

A Research group from Sudan University of Science and Technology has developed SUST-ALT database (Sudan University of Science and Technology Arabic Language Technology group) [**59][60**]. The SUST-ALT database contains numerals datasets, isolated Arabic letters datasets and Arabic names datasets. Most of these datasets are off-line. (**See Table. 2).**

Table 2: Summary of Arabic words and images databases

Name of Database	No of writers or images	Contents
CENP ARMI database [**47**]	Real life data	• 2499 words • 29498 samples of Arabic sub- words • 15175 Indian digits
IFN/ENIT database [**48**]	411 writers	• 26459 handwritten Tunisian town • 115585 piece of Arabic words (PAWs) • 212211 characters
Al-ISRA database [**55**]	500 writers	• 500 Arabic sentences • 37,000 Arabic words • 10,000 digits • 2,500 signatures
AHDB database [**56**]	100 writers	• Most commonly used Arabic words • The words are used in the writing of cheques • Handwritten pages
Arabic Printed Text Images (APTI) database [**57**]	-	• 113,284 words • 45,313,600 word images • 250 million characters
DAMP database DARPA Arabic Machine Print [**58**]	297 images text	• 60 images for development • 60 images for testing • 177 images for training
SUST- ALT database [**59**][**60**]	40.000 images text	• Numerals datasets • Isolated Arabic • Letters datasets • Arabic names datasets

6. Conclusion

This paper provides an investigation for Off-line Handwriting Arabic Text Recognition; also presented the characteristics of Arabic script. Generally, the stages of Optical Character Recognition (OCR) system is presented, focusing on initial stages, including the pre-processing and segmentation stages since they are important stages in building of OCR system. We also presented a review of the recent databases for Arabic handwriting words and images.Therefore, further research is needed in all recognition system stages especially the pre-processing and segmentation of the stages, since they are the most challenging tasks in the offline Arabic handwritten recognition system.

References

[1] Rajput, Narendrasing B., Sm Rajput, and Badave. "*Handwritten Character Recognition-A Review.*" International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8,–, 2012.
[2] N. Arica, Yarman-Vural. F, "*Optical Character Recognition for Cursive Handwriting*", IEEE Trans Pattern Anal Mach Intell, 24(6), pp.801-813, 2002.
[3] Al-Badr, B., Mahmoud, S.A., "*Survey and Bibliography of Arabic Optical Text Recognition*", Elsevier Science, Signal Processing, Vol. 41, pp. 49-77, 1995.
[4] L. Lorigo and V. Govindaraju, "*Segmentation and pre-recognition of Arabic handwriting,*" proc. Int'lconf. Document Analysis and Recognition, vol. 2, pp.605-609, 2005.
[5]Adnan Amin,"*Off-Line Arabic Character Recognition: The State of The Art*", Pattern Recognition,Vol. 31, No. 5, pp. 517-530, 1998.
[6]Jawad H AlKhateeb, Jinchang Ren, Stan S Ipson, Jianmin Jiang, "*Knowledge-based Baseline Detection and Optimal Thresholding for Words Segmentation in Efficient Pre-processing of Handwritten Arabic Text* ", Fifth International Conference on Information Technology, ,IEEE, 2008.
[7] A. M. Zeki, "*The Segmentation Problem in Arabic Character Recognition the State of The Art*", First International Conference on Information and Communication Technologies, (2005), pp. 11-26.
[8] Attia, M., "*Arabic Orthography vs. Arabic OCR; Rich Heritage Challenging AMuch Needed Technology*", Multilingual Computing & Technology magazine,USA, Dec. 2004.
[9] Al-Hajj Mohamad, R.; Likforman-Sulem, L.; Mokbel, C.; "*Combining Slanted-Frame Classifiers for Improved HMM-based Arabic Handwriting Recognition*", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No. 7, pp. 1165-1177, 2009.
[10] Khorsheed, M. S., "*Off-line Arabic character recognition – A review*", Pattern Analysis and applications, Springer-Verlag, Vol. 5, No. 1, pp. 31–45, 2002.
[11] Bdulaziz, E.Alsaif, K.I., "*Radon Transformation for Arabic Character Recognition*", International conference on Computer and Communication Engineering, Kuala Lumpur, Malaysia, pp.433-438, May 2008.
[12] S. A. Al–Qahtani, M. S. Khorsheed and M. Al–Suliman,"*Recognizing Arabic Script Using HMMs*", In Proceeding of The 17th National Computer Conference, Madinah, Saudi Arabia, King Abdulaziz University, June 2004.
[13] Santosh K.C., Cholwich Nattee. "*A comprehensive survey on on-line handwriting recognition technology and its real application to the Nepalese natural handwriting*". Kathmandu University Journal of Science, Engineering, and Technology, Kathmandu University, 2009, 5 (I), pp.31-55.
[14] M.A.A. Rashwan1, M.W.T. Fakhr2, M. Attia3 and M.S.M. El-mahallawy4, "*Arabic OCRSystem Analogous to HMM-based ASRSystems*; Implementation and Evaluation", Dec 2007.
[15] A. Hassin, H. Jian—hua, &&T. Xiang—lon."*Offline Arabic Recognition System*", Journal of Harbin Instituteof Technology (New Series), Vol. 10, No. 1, 2003.
[16] Usman Saeed. "*Automatic Recognition of Handwritten Arabic Text: A Survey* ",Life Science Journal 2014.
[17] Suliman, Azizah, Mohd S, Mohamed O, and Rahmita W. "*Chain Coding and Pre Processing Stages of Handwritten Character Image File*." electronic Journal of Computer Science and Information Technology 2, no. 1, 2011.
[18] Arica, Nafiz, Fatos T. and Yarman-Vural. "*An overview of character recognition focused on off-line handwriting*." Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 31.2 : pp.216-233, 2001.
[19] Mansi S., Gordhan B : "*A literature review on hand written character recognition*" Indian Streams Research Journal Vol -3 , ISSUE –2, ISSN:-2230-7850, March.2013.
[20] Otsu. "*A threshold selection method from gray-scale histogram*". IEEE Transactions on System", Man, and Cybernetics. 9, pp. 62-66, 1979.
[21] Zhixin S., Srirangaraj S. , Venu G. :"*Pre-processing Issues in Arabic OCR",*V. Märgner, H. El Abed (eds.),*Guide to OCR for Arabic Scripts*, Springer- Veerlag London, 2012.
[22] Ahmed T. Sahlol, Cheng Y. Suen, Mohammed R. Basyouni, Abdelhay A. Sallam "*A Proposed OCR Algorithm for the Recognition of Handwritten Arabic Characters*" Journal of Pattern Recognition and Intelligent Systems, Vol. 2 Iss. 1, pp. 90-104, Feb. 2014.

[23]Srihari, Sargur N., and Gregory Ball. "*An assessment of Arabic handwriting recognition technology.*" Guide to OCR for Arabic Scripts. Springer London, pp.3-34, 2012.
[24] Azizi, N., Farah, N., Sellami, M.: "*Ensemble classifier construction for Arabic handwritten recognition"*. Proceedings of 7th International Workshop on Systems, Signal Processing and their Applications, Algiers, Algeria, pp. 271-274, 2011.
[25] Boukerma, Hanene, and Nadir F. "*A novel Arabic baseline estimation algorithm based on sub-words treatment.*" Frontiers in Handwriting Recognition (ICFHR), International Conference on. IEEE, 2010.
[26] Parvez, M.T., Sabri, A.M.:"*Offline Arabic handwritten text recognition: A Survey*", ACM Comput. Surv. , 45, (2), pp. 1-35, 2013.
[27]Steinherz, T., Intrator, N., Rivlin E. : "*Skew Detection via Principal Components Analysis*".Document Analysis and Recognition, ICDAR '99. Proceedings of the Fifth International Conference. IEEE, 1999.
[28] Hassanin M. Al-Barhamtoshy and Mohsen A. Rashwan. "*Arabic OCR Segmented-based System*", Life Sci J;11(10), pp.1273-1283], 2014.
[29] Xiu P. Xiu, L. Peng, X. Ding, &&H. Wang ."*Offline Handwritten Arabic Character Segmentation with Probabilistic Model*", Image Processing: Algorithms and Systems XIII, , Proc. of SPIE-IS&T Vol. 9399 939911-1, 2015.
[30] Khader M; Muna A; Aziz Q; Iyad T "*Printed Arabic Optical Character Segmentation*" SPIE 9399, Image Processing: Algorithms and Systems XIII, 2015.
[31] Zeki, A. M. and Zakaria, *M. S. Challenges in Recognizing Arabic Characters.* The national conference for computer. Abu-al-Aziz king University. ArabiaSaudi .April 2004
[32] Lam, L. Lee, S.W. & Suen, C.Y. "*Thinning methodologies- A comprehensive survey,",*IEEE Trans. On Pattern Analysis Machine Intelligence. Vol. 14, No. 6, pp.869- 885,1992.
[33]Jafaar A.and Xue Li, "*An Effective Approach to Offline Arabic Handwriting Recognition*", International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 6, November 2013.
[34]Safwan Wshah, Zhixin Shi and Venu Govindaraju, "*Segmentation of Arabic Handwriting based on both Contour and Skeleton Segmentation*",2009.
[35] Manal A. Abdullah, Lulwah M. Al-Harigy, and Hanadi H. Al-Fraidi," *Off-Line Arabic Handwriting Character Recognition Using Word Segmentation*", journal of computing, volume 4, issue 3, issn 2151-9617, March 2012.
[36] Devijver, P. A., and Kittler, J., "*Pattern Recognition: A Statistical Approac*h", Englewood Cliffs, NJ: Prentice-Hall, 1982.
[37] S. Alma'adeed, "*Recognition of Off-line Handwritten Arabic Words Using Neural Network*", In Geometric Modeling and Imaging-New Trends, (2006), pp. 141–144.
[38]. K. Khatatneh, I. M. El-Emary and B. Al-Rifai, "*Probabilistic Artificial Neural Network for Recognizing The Arabic Handwritten Characters*", Journal of Computer Science, vol. 3, (2006), pp. 881–886.
[39]. A. Asiri and M. S. Khorsheed, "*Automatic Processing of Handwritten Arabic Forms Using Neural Networks*", In Proceedings Of World Academy of Science, Engineering And Technology, (2005), pp. 147–151.
[40]. A. Dehghani, F. Shabani and P. Nava, "*Off-line Recognition of Isolated Persian Handwritten Characters Using Multiple Hidden Markov Models*", International Conference on Information Technology: Coding and Computing, (2001), pp. 506–510.
[41]. S. Alma'adeed, C. Higgens and D. Elliman, "*Recognition of Off-line Handwritten Arabic words using Hidden Markov Model Approach*", In 16th International Conference on Pattern Recognition, vol. 3, (2002), pp. 481–484.
[42]. H. Al-Muhtaseb, M. Rami and S. Qahwaji, "*Recognition of Off-line Printed Arabic Text Using Hidden Markov Models*: Signal Processing, vol. 88, (2008), pp. 2902–2912.
[43]. J. H. Alkhateeb, F. Khelifi, J. Jiani and S. S. Ipson, "*A New Approach for Off-line Handwritten Arabic Word Recognition Using KNN Classifier*", In IEEE International Conference on Signal and Image Processing Applications, (2009), pp. 191–194.
[44] Ahlam MAQQOR, Akram HALLI, and Khaled SATORI, "*A MULTI-STREAM HMM APPROACH TO OFFLINE HANDWRITTEN ARABIC WORD RECOGNITION*", International Journal on Natural Language Computing (IJNLC) Vol. 2, No.4, August 2013.

[45] Khaoula ADDAKIRI , Mohamed BAHAJ "*On-line Handwritten Arabic Character Recognition using Artificial Neural Network*" International Journal of Computer Applications (0975 – 8887) Volume 55– No.13, October 2012 .
[46] Amin, A., "*Off-Line Character Recognition; A Survey*", Proceedings of the 4th Int. Conf. Document Analysis and Recognition (ICDAR '97), Ulm, Germany,pp. 596-599, August 1997.
[47] Al-Ohali, Yousef, Mohamed C., and Ching S.. " *Database for recognition of handwritten Arabic checks*." Pattern recognition 36.1: pp.111-121, 2003.
[48] El Abed, Haikal, and V. Margner. "*The IFN/ENIT-database-a tool to develop Arabic handwriting recognition systems*." Signal Processing and Its Applications. ISSPA. 9th International Symposium on. IEEE, 2007.
[49] V. Märgner, M. Pechwitz, and H. El Abed, "*ICDAR 2005 Arabic handwriting recognition competition*," in Proceedings of the 8th Inter. Conf. on Document Analysis and Recognition, vol. 1, pp. 70–74, 2005.
[50]V. Märgner and H. El Abed, "*ICDAR 2007 – Arabic Hand-writing Recognition Competition*," in Proceedings of the 9 the International Conf. on Document Analysis and Recognition (ICDAR), vol. 2, 2007.
[51] Märgner, V., & Abed, H. El.. "*ICDAR 2009 Arabic Handwriting Recognition Competition*". 2009 10th International Conference on Document Analysis and Recognition, 1383–1387, 2009.
[52] Margner, Volker, and Haikal El Abed. "*ICFHR 2010-Arabic handwriting recognition competition*." Frontiers in Handwriting Recognition (ICFHR), International Conference on. IEEE, 2010.
[53] Margner, V., & Abed, H. El. . "*ICDAR 2011 - Arabic Handwriting Recognition Competition.*" International Conference on Document Analysis and Recognition, pp1444–1448, 2011.
[54] Mezghani, Anis, Slim Kanoun, Maher Khemakhem, and Haikal El Abed. "*A database for Arabic handwritten text image recognition and writer identification.*" In Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition, pp.399-402. IEEE Computer Society, 2012.
[55] N. Kharma, M. Ahmed and R. Ward, "*A New Comperehensive Database of Hand-written Arabic Words, Numbers, and Signatures used for OCR Testing*," IEEE Canadian Conference on Electrical and Computer Engineering, vol. 2, pp. 766–768, 1999.
[56] Al-Ma'adeed, S., Elliman, D., Higgins, C.A." *A data base for Arabic handwritten text recognition research*". In Eighth International Workshop on Frontiers in Handwriting Recognition (2002).
[57] Slimane, F., Ingold, R., Kanoun, S., Alimi, A.M., Hennebert, J. "*A new Arabic printed text image database and evaluation protocols*", In: Proc. 10th International Conference on Document Analysis and Recognition, pp. 946-950, 2009.
[58] Davidson, R. B., Hopley, R. L., "*Arabic and Persian OCR training and test data sets"*, In: Proc. SDIUT, Annapolis, MD, pp.303-307 1997.
[59] Musa, M.E.M., "*Arabic handwritten datasets for pattern recognition and machine learning*," Application of Information and Communication Technologies (AICT), 5th International Conference on , vol., no., pp.1-3, 2011.
[60] Mahmoud, Sabri A., Irfan A., Mohammad A., Wasfi G. Al-Khatib, Mohammad T., Gernot A., Volker M., and Haikal E.. "KHATT: "*Arabic offline handwritten text database."* In Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition, pp.449-454. IEEE Computer Society, 2012.

