Volume 4, Issue 9, September 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Off-Line Handwriting Arabic Text Recognition: A Survey

Iman Yousif
Collage of Computer Science & Information Technology
Sudan University of Science and Technology
Khartoum, Sudan

Adnan Shaout
The University of Michigan - Dearborn
The Electrical and Computer Engineering Department
Dearborn, Michigan

Abstract- Automatic recognition of offline Handwriting in general is a difficult task. However, comparing with other language, due the nature of Arabic scripts, recognition of Arabic scrip is more difficult and present unique technical challenges. Recently, researcher's attentions in this area have increased and different methods have been applied. This paper survey the recent devolvement in the field of the offline handwritten Arabic word and text recognition and it provides a comprehensive review of these methods in each stage of the recognition system. The survey provides a comprehensive state of the art of offline handwriting Arabic Text recognition. It presents a comparison among all existing techniques with respect to various characteristics such as recognition rate, etc. The paper also presents a new proposal for an off-line handwriting Arabic text recognition system based on fuzzy decision trees.

Key Words: offline Arabic Handwriting Recognition, OCR, AI, Artificial Neural Networks, Hidden Markov Models, k-Nearest Neighbors, Fuzzy logic, Fuzzy Decision Tree.

I. INTRODUCTION

Character recognition has been one of the most fascinating and challenging research areas in the field of image processing and pattern recognition in recent years. Due to the variability in writing style and sizes, recognition of handwritten scripts is even more challenging than printed scripts. In general, Character recognition can be defined as the task of transforming text represented in spatial form of graphical marks into its symbolic representation [1] [2].

Character recognition systems are of two types, on-line and off-line systems. The main difference between them is that in an on-line system the recognition is performed at the time of writing (e.g. tablet, smart phone) while the off-line handwritten recognition is performed after the writing is completed (e.g. scanned document). Since more information is available in on-line Handwritten system, usually the recognition in on-line is easier [1] [2]. Off-line handwritten character recognition system is very important for the creation of electronic libraries, digital copies of handwritten documents and data entries. It also can provide a solution for many automated processes tasks for huge amounts of data such as automatic mail sorting, check processing, signature verification, writer identification and document analysis, etc.

Although handwriting Arabic is cursive script, most of the research in this area handles isolated characters; some researchers published papers about Arabic character recognition [3-9], some about Arabic-Indian numerals [10-19] and some included both [20]. Different method approaches have been used. Most of these methods are based on neural network, hidden Markov model and fuzzy logic.

The first commercial Optical character Recognition system (OCR) for Latin script was launched in middle of the 1950's [21]. However, due to the paucity of researches on Arabic OCR at that time, the first published paper was in 1975 and the first OCR system for Arabic was made available in the 1990s [22]. The availability of powerful inexpensive CPUs, open databases for Arabic handwritten characters, and the availability of words and text recognition researches have caused the researcher's interest to increase in this area. Several researches have been focused on new techniques and methods that would reduce the processing time while providing higher recognition accuracy. The first survey on Arabic off-line recognition area was published in 1995 [23]. Few survey papers that were dedicated to Arabic script recognition have been published [24-29] since 1995. This paper will present, to our knowledge, the first survey that focuses on handwriting Arabic words and text recognition. The rest of this paper is organized as follows: section 2 discusses some characteristics of Arabic script; section 3 describes a general model of OCR system and presents the recent research in each stage; section 4 presents different databases that are used for Arabic handwriting recognition system; section 5 presented the recent off-line Arabic word and/or text recognition systems with tables of comparisons that summarize the features, classifiers, testing data, and the recognition rates; section 6 proposed a novel off-line Arabic text recognition system based on fuzzy decision tree and finally section 7 concludes the paper.

II. ARABIC SCRIPT CHARACTERISTICS

Arabic is a widely used language, not only by Arabs in more than 23 countries in the Middle East and North Africa but is also spoken as a second language by several Asian countries in which Islam is the principle religion (e.g. Indonesia). Also there are several languages that have adopted Arabic alphabetic such as Farsi, Urdu, Malay, and some West African languages such as Hausa [12] [13].

د Dal	خ Khaa	ح Haa	ج Jeem	ث Thaa	ت Taa	ب Baa	أ Alif
ط Ttaa	ض Dahd	ص Saad	ش Sheen	س Seen	ز Zain	ر Raa	ذ Thal
م Meem	ل Lam	ك Kaaf	ق Qaaf	ف Faa	غ Ghain	ع Ain	ظ Tthaa
				ي Yaa	و Waw	هـ Hhaa	ن Noon

Figure 1: Isolated Arabic Alphabit

Final Form	Middle Form	Initial Form	Isolated Form	Final Form	Middle Form	Initial Form	Isolated Form	Final Form	Middle Form	Initial Form	Isolated Form
ق	ـقـ	قـ	ق	ـز	ـ	ـ	ز	ـا	ـ	ـ	أ
ك	ـكـ	كـ	ك	ـس	ـسـ	سـ	س	ـب	ـبـ	بـ	ب
ل	ـلـ	لـ	ل	ـش	ـشـ	شـ	ش	ـت	ـتـ	تـ	ت
م	ـمـ	مـ	م	ـص	ـصـ	صـ	ص	ـث	ـثـ	ثـ	ث
ن	ـنـ	نـ	ن	ـض	ـضـ	ضـ	ض	ـج	ـجـ	جـ	ج
ـه	ـهـ	هـ	هـ	ـط	ـطـ	طـ	ط	ـح	ـحـ	حـ	ح
و	ـ	ـ	و	ـظ	ـظـ	ظـ	ظ	ـخ	ـخـ	خـ	خ
ي	ـيـ	يـ	ي	ـع	ـعـ	عـ	ع	ـد	ـ	ـ	د
ء	ـ	ـ	ء	ـغ	ـغـ	غـ	غ	ـذ	ـ	ـ	ذ
				ـف	ـفـ	فـ	ف	ـر	ـ	ـ	ر

Figure 2: Arabic Chracters Formes

Arabic scripts are a cursive-type which are written in horizontal lines from right to left and have 28 letters as shown in figure 1. However, some additional letters are used when writing foreign words that contain sounds which do not occur in standard Arabic or when writing other languages using Arabic alphabetic (e.g. ٹ ،پ in Urdu language). Most Arabic alphabets change their form depending on the position within the word. Most alphabets have four different shapes; isolated; at the beginning; at the middle or at the end as see in figure 2. Some of Arabic alphabet strokes look exactly the same and only differ by having dots (one, two, or three) above or under the letter (e.g. ب, ت and ث). Additionally some alphabets have secondary character called Hamza (ء). This secondary character can be above the main character (like أ ك ئ and ؤ) or under the main character (like إ). Figure 3 illustrates some of Arabic script characteristics using a paragraph of Arabic handwritten.

Arabic text has small marks, called diacritics, which are used as vowels that may change the meaning of a word. For example the words (زَرَعَ)and(زَرَعْ) have the same main characters but are pronounce differently; the first word (زَرَعْ) is a noun and it means plants and the second word (زَرَعَ) is a verb which means a plant and the only difference between them is that the first word has a diacritic called sukkun (ْ) above the last character and the second has a different diacritic called fat-ha (َ). Figure 4 shows the Arabic diacritics.

The use of ligature in Arabic text is common. A ligature is defined as over-lapping combination between two characters. This combination is some time optional like in meem-haa (محـ) and laam-meem (لمـ) or not as in laam-alef that can have two forms depending on the writing style (لا ,ﻟ). The existing of the ligature makes the segmentation process more challenging. A solution to this problem is to consider ligatures as additional classes [30].

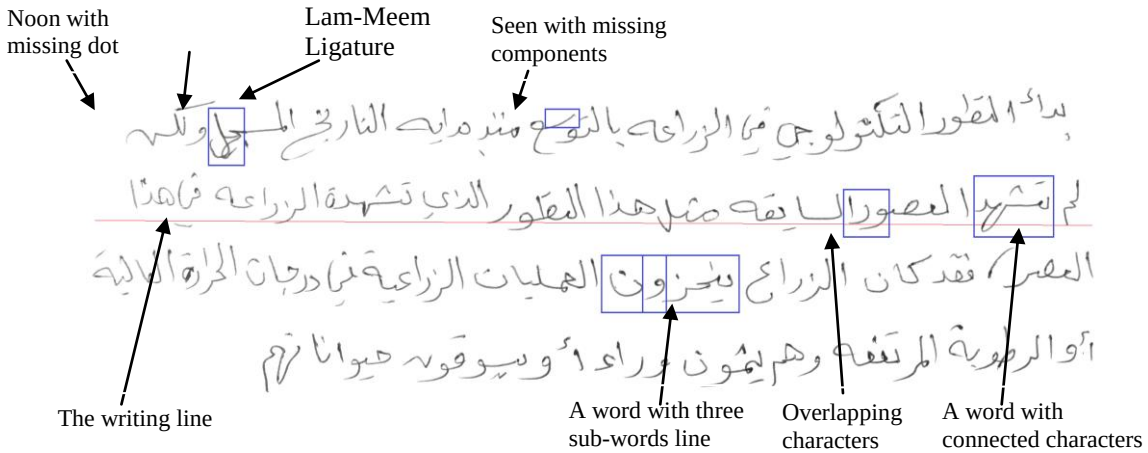

Figure 3: Some of Arabic script characteristic

fat-ha	Dhammah	Kasrah	Tanween Fat-h (double Fat-ha)	Tanween dam (double Dhammah)	Tanween kaser (double Kasrah)	sukkun	Shaddah	Maddah
◌َ	◌ُ	◌ِ	◌ً	◌ٌ	◌ٍ	◌ْ	◌ّ	~

Figure 4: Arabic diacritics

III. OPTICAL CHARACTER RECOGNITION (OCR)

The task of recognizing offline characters is called Optical Character Recognition (OCR). This name came from converting a scanned document of handwritten, typewritten or printed text using optically digitizing device such as optical scanner or camera into machine-encoded text [31]. The digital image then goes throw five major stages as shown in figure 5.

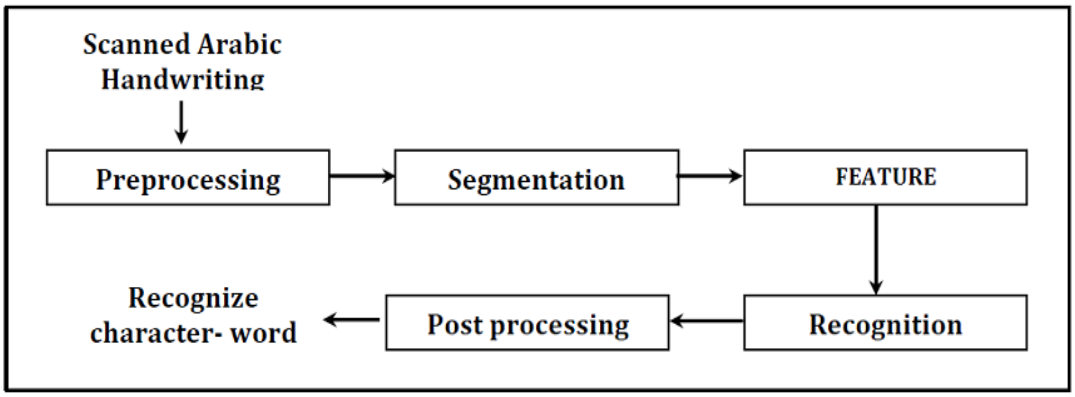

Figure 5: A typical OCR system

3.1 Preprocessing Stage

This is the first step in the character recognition system. The aim of this stage is to present an enhanced version of the original image to be more suitable for the next stage. The image will pass a number of operations like filtering, binarization, thinning, smoothing, baseline detection, skew and slant detection. Each of these operations are explained as follows [32] [33]:

- Filtering (Noise Removal) : Remove all the unwanted pixels which do not belong to the word shape.
- Binarization: Conversion of a gray scale image into binary image where a pixel can have only one of two values 0 or 1.
- Thinning: Convert the text image to some representation which is easier to process. This representation could be a skeleton which is a one-pixel thick representation showing the centerlines of the word or it could be a contour that represent the region of the text by describing its contour. The most popular method for representing the contour is the Freeman chain code [34]. Figure 6 shows an example.
- Smoothing: this operation is used to reduce the noise or to straighten the edges of the characters.
- Base line detection: this is one of the major challenges in Arabic handwriting recognition and it can affect the efficiency of the features extraction, the segmentation stage and skew normalization. The aim of this operation is to find the baseline of each word or sub-word and rotate it on its center of gravity so that the baseline becomes horizontal. For Arabic handwritten recognition several methods have been used for detecting the base line. Horizontal projection of the word skeleton in one of those methods [35, 36]. Figure 7 shows a sample. Hanene Boukerma and Nadir Farah develop baseline estimation algorithm based on Sub-Words [37] because Arabic words contain often more than one part of Arabic word (PAW), and some time those PAWs have different slant angles within the same word.

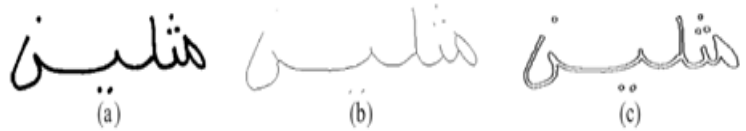

Figure (6): (a) original word, (b) word skeleton, (c) word contour.

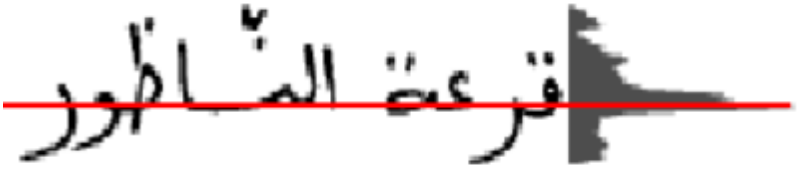

Figure (7) : Horizontal Projection Method for Detecting Arabic Baseline.

3.2 Segmentation Stage

After the preprocessing stage, the text image may need to be dividing into objects. It splits the paragraph into separate lines and then split these lines into words or sub words then to characters or sub characters to be recognized. Segmentation is considered to be one of the most important and challenging tasks in the OCR system. The impact of this stage will affect the overall system performance.

The cursive nature of the Arabic script, the overlapping between the characters, different forms of each letter depending on its position in the word or the writing style and the presence of the secondary characters like dots, Hamza and diacritics, are all factors that increase the difficulty of this stage [38][39]. During the past few years there have been promising attempts by researchers to solve this problem; some of this work is summarized and compared in Table 1.

Dinges et al. [40] presented a novel locale grouping based method for line segmentation of handwritten Arabic documents. After using Support Vector Machine (SVM) to classify the entire connected component as PAW or diacritics, they used their developed distance measures to calculate the nearest neighbors between all PAWs. The next step they used was a graph based grouping algorithm to build all the candidate lines starting from the first PAW from the right for all lines. Finally, all the unused PAWs are assigned to the nearest line. There method works fine with documents of different writers and styles even if the text lines have unequal skews or curvatures.

To makes use of the nature of Arabic script, Eraqi and Abdelazeem [41] combined the local writing direction information and the neighborhood geometric characteristics to propose a new efficient explicit technique for segmentation of offline Arabic handwriting which segmented the text into basic graphemes. The proposed technique applies the Douglas-Peucker algorithm on the skeletonized parts of the offline handwriting images. This method has proved effective; as it obtained 91.27% of the correctly segmented graphemes using 1000 images containing 1402 Arabic handwritten words and 7960 Arabic handwritten graphemes taken from the IFN/ENIT database.

Lawgali et al. [42] exploited the fact that segmentation points, which occur at the end of a character and the beginning of the next, are usually located in the region surrounding the baseline to present a segmentation algorithm of Arabic handwritten words. The segmentation algorithm starts with segmenting the word into sub-words and then the baseline of each sub-word is computed. The algorithm then deletes all the descended sub-words which have a starting point below the baseline. The vertical projection is used to find the candidate points for the segmentation. The algorithm has been tested using 800 handwritten Arabic words taken from IFN/ENIT database and has achieved 82.98% character accuracy. However, this algorithm couldn't segment the alphabets (س, ش) into three segments rather it only segmented them into one.

Al-Khateeb et al. [43] introduced a words segmenting methods for Arabic handwritten text. After extracting the connected components (CCS) and distances among different components are analyzed then the statistical distribution of this distance is obtained to determine an optimal threshold for word segmentation. Meanwhile, an improved projection based method is also employed for baseline detection. The proposed method has been tested with 200 images from the IFN/ENIT database and has obtained 85% accuracy.

Al Hamad and Abu Zitar [39] presented a three steps segmentation method. The first step is based on feature-based Arabic Heuristic Segmentor (AHS) to obtain over-segmentation from the thinned words. Step two applied Neural-based segmentation point on those initial segmentation points to validate them. Finally, the outputs of the previous networks are then combined to decide whether a particular segmentation point is valid or not. The segmentation achieved accuracy of 82.98% on 500 words written by ten writers.

Al Hamad [44] proposed fusion equations for improving the segmentation of word image. This method has two phases. In the first phase the author applied AHS to place the Prospective Segmentation Points (PSP) in the whole parts of the word image. In the second phase the author applies Neural-based segmentation technique to examine all PSPs and identify the invalid ones. This method has been implemented and tested on 425 word image from local benchmark database and has achieved 88.96% accuracy.

Tamen and Drias [45] tried to overcome the over-segmentation problem in the segmentation stage by pasting the segmented parts to rebuild the whole character form after the rejection or the ambiguousness decision in the recognition stage. First, they used multilayer perceptrons (MLP) in the recognition process but in order to improve the system performance, they replaced the MLP by a linear feed forward network. The training was done using the back propagation algorithm with all the pre segmented Arabic characters and their different positions written by three different persons. For testing the system the authors used texts written by three other persons.

Parvez & Mahmoud [46] presented a robust lexicon reduction segmentation algorithm to segment Arabic words into graphemes. This method is based on the characteristic of Arabic script; which indicates predictable segmentations of Arabic characters. The authors tested there method on 32,492 images from the IfN/ENIT database.

Osman [38] developed a segmentation algorithm for Arabic handwriting. The first step in the algorithm was to divide the selected image into lines and sub-words, then trace the sub-word contour. Finally, the algorithm detects the exact points where the contour changes its state from a horizontal to vertical or curved line and consider those point as a segmentation points. The algorithm achieved 89.4% segmentation accuracy on 537 tested words from the IFN/ENIT database.

Samoud et al. [47] presented two combining methods for segmenting Arabic handwritten script into characters. The first method was based on the analysis of the contour (Min-Max) minima and maxima and the projection. The second method was based on Hough Transform (HT) and Mathematical Morphology (MM) operators. To compare the two methods the authors used three evaluation criteria's; segment positions (SP), segment numbers (SN) and the recognition rates. For the two methods, the segmentation rate was less than 30% on a data set from the IFN/ENIT Database.

Table 1: Comparison between some segmentation methods

Author	Year	Segmentation Scope	Test Data	Segmentation Method	Accuracy
Dinges et al.	2013	Line segmentation	-	Grouping based method	-
Eraqi and Abdelazeem	2012	Graphemes	1402 words		91.27%
Lawgali et al.	2011	Characters	800 words	Extracting baseline	82.98%
Al-Khateeb et al.	2008	Words	200 images	Component-based method	85%
AlHamad and Abu Zitar	2010	Characters	500 words	Over-segmentation & ANN	82.98%
Tamen and Drias	2010	Characters	Texts written by three persons	Multilayer perceptrons then ,back propagation	Unknown
Parvez & Mahmoud	2013	Graphemes	32,492 images	Robust lexicon reduction	
Al Hamad	2013	Characters	425 word images	AHS and Neural	88.96%
Osman	2013	Lines, sub-words and characters	537 words	Contour extracting points	89.4%
Samoud et al	2012	Characters	1250 images	Min-Max-projection HT-MM	Less than 30%

3.3 Feature Extraction Stage

This is also an important stage in the OCR system. Feature extraction is the process of getting useful information from the word/character image. The information will be used to generate modules to train the classifier and to be used for classification purposes [48]. In general there are two categories of features extracted, structural and statistical features. Choosing the wright feature extraction method might be the most important step for achieving a high recognition rate [49]. However, in some cases the combination of several features extraction types could be a wise decision to enhance the overall recognition performance.

Structural features are the character/word image geometrical and topological information. Those obtained information include the number of PAWS, descenders, ascenders, dot below the baseline, above the baseline, etc. Figure 8 shows a structural features example. Statistical feature are numerical measures computed over the images. They include pixel densities, histograms of chain code directions, moments, Fourier descriptors, etc. [50] [51]. In HMM classifier based system, it is usual to use sliding windows for extracting features from the word image [52].

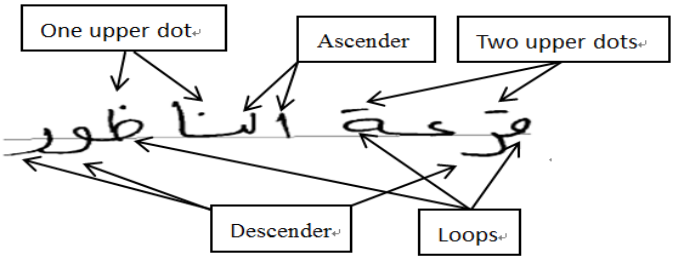

Figure 8: Structural Features of Tunisian Town Name قرعة الناطور that contain Two words and seven Paws

3.4 Training and Recognition (Classifications) Stage

This stage is considered as the primary stage for the OCR system. It depends on the previous stages so that defect in the earlier stages will affects the recognition process and will lead to a low recognition rate. More information about this stage will be covered in the classifications methodology section.

3.5 Post processing Stage

This is the final stage in the OCR system. This stage can improve recognition accuracy and the system performance by refining the decisions taken by the previous stage and possibly recognizing words by using the context [1] [46].

IV. ARABIC HANDWRITING WORDS AND TEXT DATABASE

With the increasing interest in Arabic handwriting recognition, the need for a freely standard Arabic handwriting database that represents variety of handwriting styles is highly required. In the past, the lack of freely available Arabic databases is considered as one of the reasons for the lack of research on Arabic text recognition compared with other languages. Most of research groups implemented their system on set of data gathered individually. Therefore, the comparison of OCR Arabic systems was not reasonable in the past. Currently, there are several Arabic text and words databases to serve handwritten Arabic characters, digit, word and text recognition research. Table 2 illustrates some of Arabic text and words databases.

Al-Ohali et al. [53] in the Centre for Pattern Recognition and Machine Intelligence (CENPARMI) in Montréal has developed a Database that can be used by researcher in the field of Arabic handwritten Arabic legal amounts recognition. What most distinguishes this database is the data are extracted from real-life cheques collected from a financial institution which make recognition systems more adjustable to real-world applications. The database contain 29498 samples of Arabic sub-words within the domain of legal amount, 15175 samples for Indian digits and 2499 samples of each of legal and courtesy amounts written in Indian digits. CENPARMI database is divided into training and testing sets. The training set includes 66–75% of the available data. The division between training and testing data was done randomly.

Another database than can be used for legal amounts recognition researches is the AHDB (Arabic Handwritten Database) that was built by Al-Ma'adeed et al. [54]. This database were written by 100 writers and contains words and sentences that were used in writing checks. It has the most popular written Arabic words and it contains free handwriting pages in any area of writer interest.

The IFN/ENIT database of handwritten Tunisian town names [55] is the most commonly used databases by researchers who are working on Arabic handwritten recognition systems. This database was developed by the Institute of Communications Technology (IfN) at Technical University Braunschweig in Germany and the Ecole Nationale d'Inge'nieurs de Tunis (ENIT) in Tunisia. Version 1.0 of the IFN/ENIT-database consists of 26459 handwritten Tunisian town/village names, 115585 pieces of Arabic words (PAWs), and 212211 characters. Each handwritten town name comes with binary image bitmap and additional GT information. Several competitions in the past few years have been conducted using this database [56, 57, 58, 58 and 60]. Also, most of the research that was published recently have used this database.

Mezghani et al. [61] introduced an Arabic handwritten text images database written by multiple writers (AHTID/MW). The AHTID/MW contains 3710 text lines and 22896 word images written by 53 writers.

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) [62]. The SUST-ALT database contains numerals datasets, isolated Arabic letters datasets and Arabic names datasets. Most of these datasets are off-line.

A research group from King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia has developed Arabic Offline Handwritten Text Database KHATT (KFUPM Handwritten Arabic Text) [63, 64]. The database were written by 1000 different writers from different countries, gender, age groups, and handedness and education level. The database contains 2000 similar-text paragraph images and 2000 unique-text paragraph images and their extracted text line images.

Lawgali et al. [65] has developed a new database for handwritten Arabic characters (HACDB). Although this is a characters database but it can be used for training and testing words recognition after the segmentation stage because it cover all shapes of Arabic characters including overlapping ones. The HACDB contains 6600 shapes of characters written by 50 writers.

Developing a database for offline Arabic handwritten text is expensive in term of manpower and time. Therefore, Dinges et al. [66] developed an efficient system that automatically generates images of synthetic handwritten words or text lines from Unicode. The system is based on an Active Shape Models (ASM) that used online sample to generate unique letter representations for any chosen synthesis. These representations are modified by affine transformations, smoothed by B-Spline interpolation and composed to text. This system can be used as alternative to off-line handwritten samples, with variations in shape and texture.

Table 2: Samurais of Arabic Text and Words Databases

Database name	No of writers	Contents
CENPARMI - Database for handwritten Arabic checks [53]	Real-life data	• 2499 words • 29498 sub-words • 15175 Indian digits
AHDB [54]	100 writers	• Words and sentences that used in in writing checks. • popular Arabic words • Free handwriting pages.
IFN/ENIT database[55]	411 writers	• 26459 handwritten Tunisian town/village names • 115585 PAWs • 212211 characters
AHTID/MW [61]	53 writers	• 3710 text lines • 22896 word images
SUST-ALT database[62]		• numerals datasets • isolated Arabic • letters datasets • Arabic names datasets
KHATT database[63] [64]	1000 writers	• 2000 similar-text paragraph images and their extracted text line images. • 2000 unique-text paragraph images and their extracted text line images.
HACDB[65]	50 writers	• 6600 segmented characters

V. CLASSIFICATIONS METHODOLOGY

Compared to Latin and Chinese script where a lot of research work was done, the number of published work on Arabic script is quite limited. Although the number of papers published in the past few years is increasing and different techniques and methods which were intended to reduce the processing time while providing higher recognition accuracy are reported. Most of those classifying methods are based on Artificial Neural Networks (ANN), Hidden Markov Models (HMM), k-Nearest Neighbors (k-NN), Fuzzy logic (FL), Hybrid approaches and others. In general there are two basic strategies for recognizing words; Holistic strategies (Global) or Analytic strategies. The Holistic (Global) strategy recognizes the whole words or sub words without requiring segmentation, but it works on a limited number of vocabularies. The Analytic Strategy recognizes the segmented features, requires segmentation, and can be applied on unlimited vocabularies [33]. The rest of this section is categorized according to these strategies and some related works will be illustrated.

5.1 Holistic Strategies

Literal Amounts recognition is one of the most important applications in offline handwriting recognition area. Few decades ago processing checks without human involvement was just a dream. Since Literal Amounts contains a limited lexicon (48 words that can be written in an Arabic literal check amount) it seems reasonable that all the researchers in this area have used the holistic approach.

In recent years, some promising recognition systems using holistic approach have been published. Table 3 shows a summary of those attempts. In 2004 Farah et al. [67] have presented an offline Arabic check literal amount recognition system. The system is based on three parallel classifiers (ANN multilayer perceptron, k-Nearest Neighbors, and a fuzzy KNN). The input to all the classifiers is the same set of structural features. The results of the three classifiers are combined using a statistical decision system. The system obtained 96% recognition rates using a database containing 4800 words which represents the 48 words of lexicon written by 100 different writers of which 1200 wards were used for training and the rest were used for testing. Although the achieved recognition rate is satisfying, the system is not suitable for a large vocabulary lexicon.

In 2005 Farah with different group [68] has used the same database to present another Arabic literal amount recognition system. But this time, they used structural and statistical feature extraction and three parallel neural network classifiers (Multi-Layer Perceptron (MLP)). The obtained results were then combined to produce a final decision. This time the best recognition rate was 93.00%. In the same year the same group [69] produced another system that used parallel neural network classifiers feed by structural and the statistical feature extraction, but, this time one of the MLP was used as a Meta classifier. The same database has been use (2400 words for training and the 2400 for testing). Different parallel combination schemes were presented, and the best recognition rate was 95.2%.

Souici-Meslati & Sellami [70] have presented a hybrid neuro–symbolic classifier approach for recognition Arabic literal amount. The knowledge base constructed using features extracted from 48 words and they used a translation algorithm to convert the rules representation into a neural network. This system obtained 93% recognition rate. In order to evaluate this system, the authors have used 576 words written by four different writers for training and 1200 words written by 25 different writers for testing.

Based on fuzzy proximity measure especially in bank checks area L Farah et al. [71] presented another literal amount recognition system using fuzzy classifier to allocate a class to the test word on a basis of a training set. The fuzzification was introduced in two stages. The first stage was to reclassify the obtained K nearest neighbors (KNN) by a crisp KNN approach. The second stage in the classification of the tested word was to allocate it to a class among its K neighbors. The system obtaining of 93.80% recognition rate using 1200 images of 48 words written by 25 different writers.

Automatic recognition of city names and addresses in large quantity of mail is highly essential. Although postal addresses have a large vocabulary compared to literal amount lexicon. However, Holistic approach is also wildly used in postal addresses recognition.

Based on decision tree classifier, Amrouch et al. [72] has presented an offline handwritten Algerian city names recognition system. The authors have used structural features (sub words, ascenders, descenders, loops and diacritical dots) of the word images as an input for the decision tree. The system achieved 75.74% recognition rate using database contains 48 city name written three times by 100 writers for learning and testing the system.

Souci et al. [73] have presented an Arabic postal code recognition system. The system is a knowledge based artificial neural network. The first step in this system is to localize the city name from the envelope and segments it into words, then structural feature are extracted from the word contour. The knowledge base rule sets were constructed using a description of the words features. The rules are then translated by spatial algorithm for the neural network, which is trained in 550 words of 55 Algerian city names written by ten different writers. Using the same training set, a comparison was carried out by the authors between their proposed system and a MLP classifier system. The comparison showed that the training took about 10 times less than the MLP classifier and the best recognition rate achieved was 92%.

Table 3: Summary of results for Literal Amounts and City Name Recognition Systems

authors	Year	Representation	Feature	Classification methods	Training data	Testing data	Recognition Rates
Farah et al.[1]	2004	Structural contour	structural	ANN, KNN & Fuzzy	1200 word images	3600 word images	96%

				KNN			
Farah et al.[2]	2005	contour & diacritical dots	structural & statistical	Three Multi-Layer Perceptron (MLP) ANN ANN Multi classifiers	2400	2400	93.00%
Farah et al.[3]	2005	contour & diacritical dots	structural & statistical	ANN Multi classifiers	2400	2400	95.2%
Souici-Meslati & Sellami[4]	2004	Contour	structural	neuro–symbolic classifier	576 words written three times by four writers	1200 48 words written by 25	93%
L Farah et al[5]	2006	Contour	structural	Fuzzy K-NN	-	1200 words	93.80%.
Amrouch et al. [72]	2011	contour	structural	decision tree	14.400 words	14.400 words	75.74%
Souci et al. [73]	2004	contour	structural	knowledge based artificial neural network	550 words	550 words	92%.

5.2 Analytic Strategies

This Segmentation base recognition method is suitable for large vocabulary recognition system. Segmenting the word/sub-word into characters is required. Analytic strategies are divided into two categories, implicit and explicit base segmentation. In the Implicit based segmentation the segmentation and recognition of characters are achieved at the same time. The system searches the image for components that match the predefined classes were in explicit base segmentation the segments are identified based on "character like" properties [74] [75].

5.2.1 Hidden Markov Models Approach

The success of using Hidden Markov Models (HMMs) methods in automatic speech recognition encouraged researcher to use it in hand written recognition [76]. HMM is considered as one of the most commonly and successfully used method in offline Arabic handwritten word recognition [52] [76-84].

Based on combined scheme of HMMs and re-ranking, Al Khateeb et al. [85] have presented an Offline Arabic handwritten text recognition system. The proposed system has three main stages; preprocessing, feature extraction and classification. The features were extracted from the segmented words using sliding window. The extracted features are fed to the HMM classifier. In order to improve accuracy, the HMM result is further refined by using a re-ranking Scheme. Using the IFN/ENIT database, the system has achieved 95.15% recognition rate.

Using an explicit segmentation module, Elzobi et al. [86] have presented an off-line Handwriting Arabic words recognition system based on Hidden Markov Model. Instead of using sliding window based features; they used shape representative features for each letter in each handwritten form. They have used two databases; the IESK-arDB for training and testing, and the IFN/ENIT database samples for validation. The recognition rate have reached 71% on the first database and only 42% for the second (due to the variability in IFN/ENIT which is higher than that of IESK-arDB).

5.2.2 Artificial Neural Network Approach (ANN)

Neural network approach has performed successfully in many fields and off-line handwritten recognition is one of them. The ability to be trained automatically from examples, have faster development times , possible run on parallel processors and achieving good performance with noisy data, makes the use of ANN as a classifier appealing [87-94].

Farah [7] has implemented a neural network based system that used cascaded networks to recognize Arabic segmented characters after resizing the character image to 48 X 32 pixels block, filtration, and converting to binary in the preprocessing stage. The segmented character is further divided to 6 by 4 segments. Five features are extracted from each of 24 blocks. Those 120 features of the character are being passed to the neural network input as a single column. The basic structure of the neural networks consists of one MLP network and three LVQ networks. The data features of the image is being inputted to the MLP network. In order to minimize the complexity of the network the similar characters are being recognized as the same (ف, ي, ن, ث, ت, ب, ق). The output of first neural network will be the input of the LVQ networks after dividing it into three categories. LVQ networks have the ability to recognize very close features with lower processing time. The data set that was used to test and measure the proposed system performance consisted of 100 different separated characters that have been written by 10 different persons with the Arabic characters "Roq'a" style, the most common Arabic writing style. The recognition rate of this system was between 51% and 77% based on the character shape.

5.2.3 Fuzzy logic Approach

Using fuzzy logic in Arabic handwritten recognition seems to be very logical. The script and the variability of the Arabic script makes automatic Arabic recognition a very challenging task. A fuzzy set is similar to a classical set except that in a classical set data can either belong to the set or not whereas in a fuzzy set the data will always belong to the set but with a different degree. The degree of belonging to a fuzzy set is called a Membership [95] [96].

In 1994 Abuhaiba et al. [97] have presented an automatic off-line character recognition system for handwritten cursive Arabic characters. The system is divided into two stages, preprocessing and recognition stage. In the preprocessing stage, the first step was to skeletonize the segmented character using clustering-based skeletonization algorithm (CBSA) then the character skeleton is convert to a tree structure for recognition. In the recognition stage, a set of fuzzy constrained character graph models (FCCGM's) was designed. For recognition, a set of rules was applied to match a character tree structure to an FCCGM. The system achieved 73.6% recognition rate with 420 characters used for learning and 330 for testing.

In order to show the importance of the intuitionistic fuzzy similarity measures (IFSM), Baccour et al. [98] have applied the IFSM on a data set from the IFN/ENIT database. After extracting the features from the word image, these features were fuzzified and represented by intuitionistic fuzzy sets. IFSMs then were applied to make the comparison between the test data set which was made of 4357 word images and the training data set which was made of 2180 word images. The best obtained recognition rate was 90.78%.

Parvez and Mahmoud [20] have presented *a novel method* for recognizing isolated Arabic handwritten alphanumeric characters. After the preprocessing stage, the contour of the character image was extracted and polygonal approximation of the contour was constructed. The nearest neighbor (NN) classifier based on fuzzy attributed turning function (FATF) was used for classification. For testing and system performance the authors have used two different databases, one for handwritten Arabic characters and other for Arabic numerals. The system obtained a recognition rates of around 98% for Arabic characters and more than 97% for Arabic numerals. Then in 2013 the same authors [30] have extend their work to present the first integrated offline Arabic handwritten text recognition system based on structural techniques. In addition, they introduced several novel ideas and techniques that can be used for structural recognition of Arabic handwriting. The first step in this system was to extract the PAWs from the text lines, then perform a novel slant corrected algorithm at the PAW level. A novel segmentation algorithm was then used which was integrated into the recognition phase. The PAW were segmented into smaller components were these components may be valid Arabic characters or parts of Arabic characters. After that the best segmentation of the PAW and its constituent characters were identified by an adaptive algorithm. Multiple hypotheses were also generated for each PAW and passed through post-processing steps, like lexicon consultation, to re-rank the hypotheses and select the best matching word. In the training phase, a modeling of Arabic isolated characters was done by polygonal approximation of the characters contours. The resulting models, called Fuzzy Attributed Turning Functions (FATF). The authors compared there system with other systems using the IfN/ENIT database and achieve 79.58% recognition rate.

The work in [30] was extended by Mahmoud et al. [64] where they developed an open vocabulary offline handwritten Arabic text structural recognition system. In the recognition stage, the basic shape of the PAW without dots did pass through two levels. The first level was used to generate hypotheses for the PAW image then the segmented part from the PAW was matched with the characters model using a fuzzy distance measure. The second level generated hypotheses of the paw which was verified to leave only the best hypotheses from the first level. Finally PAW dot information wear incorporated to generate the final PAW hypotheses. The open vocabulary offline handwritten Arabic text structural recognition system was tested using 7900 isolated characters written by 52 writers. The system achieved 51.5% recognition rate using KHATT database.

5.2.4 Hybrid approach

Leila et al. [99] have presented an off-line Multiple Classifier System (MCS) to solve Arabic cursive word recognition problem. This system has two different classifiers, the Fuzzy Adaptive Resonance Theory (Fuzzy ART network) which was used for the first time in Arabic OCR, and the Radial Basis Functions (RBF). Using IFN/ENIT database the combined system had a recognition rate of 90.1 %.

Nemouchi, et al. [100] have produced a multi classifiers system for Arabic handwritten words recognition. The proposed system focused on two phases, the feature extraction and classification phases. In this system the words were represented using three feature extraction methods. The Zernike moments were extracted from binary image, the Freeman chain code was extracted from the image contour, and zoning was done on the image skeleton. Those extracted feature were used as inputs to the four parallel classifiers; Fuzzy C-Means algorithm (FCM), K-Means algorithm, K Nearest Neighbor algorithm (KNN) and a Probabilistic Neural Network (PNN)). When using all features the system obtained 80% recognition rate on 1440 words images from the Algerian city-name images database.

Farsi language uses Arabic alphabetic for writing. Therefore it seems reasonable to mention the researches that addictive to this language. Based on fuzzy vector quantization (FVQ) and hidden Markov model (HMM) Dehghan et al. [101] have presented a postal address recognition system. The proposed system was tested using 17,000 images of 198 Farsi city names with the best recognition rate of 96.5%.

VI. A NEW PROPOSED FUZZY DECISION TREE METHOD

Decision trees are considered a powerful solution structures for many applications like pattern recognition, machine learning and data mining [102]. They are capable of breaking down complex decisions into simpler decisions that can be

managed making them suitable for classification problems [103] [104]. Decision trees have been used once for off-line Arabic word handwritten recognition by Amrouch et al. [72]. They were also used for printed Arabic text recognition by A. Amin [105] and by Abuhaiba [106] for Arabic printed font recognition. Decision tree was also used for handwritten Urdu and Bangla characters recognition [107] [108].

Recently, fuzzy set theory has been combined with decision trees to produce a powerful tool that can deal with ambiguity and vagueness in real life problem. This combination is known as fuzzy decision tree which was firstly introduced by Chang and Pavlidis [109] in 1977. Since then Fuzzy decision trees have played important roles in many fields such as pattern recognition and classification. Gaolin Fang et al. [110] have used fuzzy decision tree with heterogeneous classifiers to develop Large Vocabulary Sign Language Recognition system. Kasim et al. [111] have used fuzzy decision tree to develop image classifier for Batik, one Indonesian cultural heritage image classification. Decision tree have been used in Speech Recognition [112] and in the medical field for diagnosing breast cancer [113].

Despite that success in those areas, for unknown reason fuzzy decision trees have never been used for Arabic handwritten recognition. We think fuzzy decision tree will play a major role and will achieve essential results in recognizing Arabic handwritten. Therefore, we are planning to develop an off-line Arabic handwritten text recognition system based on fuzzy decision trees. We are planning to use the IFN/ENIT database of handwritten Tunisian town names, which is consider to be the most commonly used databases by off-line Arabic recognition researchers for training and testing.

VII. CONCLUSIONS

This paper provides a comprehensive state of the art presentation of offline handwriting Arabic word and text recognition. The paper presented the unique characteristics of handwritten Arabic text and word. It also presented a survey of the recent development in the field of the offline handwritten recognition and provided a comprehensive review of these methods in each stage of recognition system and surveyed the existing Arabic word and text database.

Although Arabic language is a cursive written language, yet most of the research in literature was directed to isolated character recognition and relatively few for word and text recognition. Therefore, it is clear that offline recognition of Arabic text is still an open issue. There is still urgent need for a high speed recognition rate systems. The improvements in any stage of recognition system will lead to increasing of the overall system efficiency. Therefore, more research is needed in all the recognition system stages especially the segmentation and the classification stages, since they are the most challenging tasks in the offline Arabic handwritten recognition system.

New and improved offline Arabic handwritten recognition systems can be generated through extracting different kind of features, combining between different technologies or experiencing techniques that have never been used before. We believe that the use of fuzzy decision trees could lead to a remarkable new offline Arabic handwritten recognition system.

REFERENCE

[1] Ahmed, Pervez, and Yousef Al-Ohali. "Arabic character recognition: Progress and challenges." Journal of King Saud University-Computer and Information Sciences 12 (2000): 85-116.

[2] AL-Shatnawi, Atallah Mahmoud, AL-Salaimeh, Safwan, AL-Zawaideh, Farah Hanna, Omar, Khairuddin, 2011. Offline Arabic text recognition an overview. World Computer. Sci. Inform. Technol. J. 1 (5), 184–192

[3] Ali, Mohamed A. "A classifier for Arabic handwritten characters based on supervised self-organizing map neural network." Proceedings of the 2010 international conference on Mathematical models for engineering science. World Scientific and Engineering Academy and Society (WSEAS), 2010.

[4] Abed, Majida Ali, H. Abed, Z. Baha, and A. Ismail. "Fuzzy Logic approach to Recognition of Isolated Arabic Characters." Int. Jour. of Computer Theory and Engineering 2, no. 1 (2010): 119-124.

[5] Elglaly, Yasmine, and Francis Quek. "Isolated Handwritten Arabic Character Recognition using Multilayer Perceptrons and K Nearest Neighbor Classifiers." (2011).

[6] Gharehchopogh, Farhad Soleimanian, and Ezzat Ahmadzadeh. "Artificial Neural Network Application in Letters Recognition for Farsi/Arabic Manuscripts." International Journal of Scientific & Technology Research 1.8 (2012): 90-94.

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

[8] Dinges, Laslo, Ayoub Al-Hamadi, and Moftah Elzobi. "An Active Shape Model based approach for Arabic handwritten character recognition." Signal Processing (ICSP), 2012 IEEE 11th International Conference on. Vol. 2. IEEE, 2012.

[9] Sahlol, Ahmed, and Cheng Suen. "A Novel Method for the Recognition of Isolated Handwritten Arabic Characters." arXiv preprint arXiv: 1402.6650 (2014).

[10] Mahmoud, Sabri A., and Sameh M. Awaida. "RECOGNITION OF OFF-LINE HANDWRITTEN ARABIC (INDIAN) NUMERALS USING MULTI-SCALE FEATURES AND SUPPORT VECTOR MACHINES VS. HIDDEN MARKOV MODELS." Arabian Journal for Science & Engineering (Springer Science & Business Media BV) 34 (2009).

[11] Montazer, Gholam Ali, Hamed Qahri Saremi, and Vahid Khatibi. "A neuro-fuzzy inference engine for Farsi numeral characters recognition." Expert Systems with Applications 37.9 (2010): 6327-6337.

[12] Mahmoud, Sabri A., and Sunday Olusanya Olatunji. "Handwritten Arabic numerals recognition using multi-span features & Support Vector Machines." Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on. IEEE, 2010.

[13] Mahmoud, Sabri A., and Marwan H. Abu-Amara. "Recognition of handwritten Arabic (Indian) numerals using Radon-Fourier-based features." Proceedings of the 9th WSEAS International Conference on Signal Processing, Robotics and Automation, (ISPRA'10), ACM Press, USA. 2010.

[14] Lawal, Isah A., Radwan E. Abdel-Aal, and Sabri A. Mahmoud. "Recognition of handwritten Arabic (Indian) numerals using freeman's chain codes and abdicative network classifiers." Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, 2010.

[15] ALI, ABDULBARI AHMED, and RJ RAMTEKE. "FUZZY BASED RECOGNITION OF HANDWRITTEN ARABIC NUMERALS." International Journal of Machine Intelligence 3.3 (2011).

[16] Azeem, Sherif Abdel, and Maha El Meseery. "Arabic Handwriting Recognition Using Concavity Features and Classifier Fusion." Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on. Vol. 1. IEEE, 2011.

[17] Singh, Pratibha, Ajay Verma, and Narendra S. Chaudhari. "Classification of Hindi numeral using Fuzzy Zoning and SVM." Advanced computer and communication conference. 2011.

[18] Zaghloul, Rawan I., Dojanah MK Bader Enas, and F. AlRawashdeh. "RECOGNITION OF HINDI (ARABIC) HANDWRITTEN NUMERALS." American Journal of Engineering and Applied Sciences 5.2 (2012): 132.

[19] AlKhateeb, Jawad H., and Marwan Alseid. "DBN-Based learning for Arabic handwritten digit recognition using DCT features." Computer Science and Information Technology (CSIT), 2014 6th International Conference on. IEEE, 2014.

[20] M. T. Parvez and S. Mahmoud, "Arabic Handwritten Alphanumeric Character Recognition using Fuzzy Attributed Turning Functions," in First International Workshop on Frontiers in Arabic Handwriting Recognition, 2011.

[21] Eikvil, Line. "Optical Character Recognition." citeseer.ist.psu. Edu/142042. Html (1993).

[22] Märgner, Volker, and Haikal El Abed. "Arabic word and text recognition-current developments." Proceedings of the Second International Conference on Arabic Language Resources and Tools, Cairo, Egypt, April. The MEDAR Consortium. 2009.

[23] Al-Badr, Badr, and Sabri A. Mahmoud. "Survey and bibliography of Arabic optical text recognition." Signal processing 41.1 (1995): 49-77.

[24] Amin, Adnan. "Off line Arabic character recognition: a survey." Document Analysis and Recognition, 1997. Proceedings of the Fourth International Conference on. Vol. 2. IEEE, 1997.

[25] Amin, Adnan. "Off-line Arabic character recognition: the state of the art." Pattern recognition 31.5 (1998): 517-530.

[26] Khorsheed, Mohammad S. "Off-line Arabic character recognition–a review." Pattern analysis & applications 5.1 (2002): 31-45.

[27] Lorigo, Liana M., and Venugopal Govindaraju. "Offline Arabic handwriting recognition: a survey." Pattern Analysis and Machine Intelligence, IEEE Transactions on 28.5 (2006): 712-724.

[28] Jumari, Kasmiran, and Mohamed A Ali. "A survey and comparative evaluation of selected off-line Arabic handwritten character recognition systems." Jurnal Teknologi 36.1 (2012): 1-18.

[29] Parvez, Mohammad Tanvir, and Sabri A. Mahmoud. "Offline Arabic handwritten text recognition: a survey." ACM Computing Surveys (CSUR) 45.2 (2013): 23.

[30] Tanvir Parvez, M., & Mahmoud, S. a. (2013). Arabic handwriting recognition using structural and syntactic pattern attributes. *Pattern Recognition*, *46*(1), 141–154. doi:10.1016/j.patcog.2012.07.012

[31] RAJPUT, MR NARENDRASING B., SM RAJPUT, and SM BADAVE. "Handwritten Character Recognition-A Review." International Journal of Engineering Research & Technology (IJERT) Vol. 1 Issue 8, October – 2012 ISSN: 2278-0181

[32] Suliman, Azizah, Mohd Nasir Sulaiman, Mohamed Othman, and Rahmita Wirza. "Chain Coding and Pre Processing Stages of Handwritten Character Image File." electronic Journal of Computer Science and Information Technology 2, no. 1 (2011).

[33] Arica, Nafiz, and Fatos T. 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 (2001): 216-233.

[34] Al-rashaideh, H. (2006). Preprocessing phase for Arabic Word Handwritten Recognition. Information Transmissions in Computer Networks Journal, 6(1), 11–19. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.95.5611

[35] Srihari, Sargur N., and Gregory Ball. "An assessment of Arabic handwriting recognition technology." Guide to OCR for Arabic Scripts. Springer London, 2012. 3-34.

[36] Atallah, AL-Shatnawi, and Khairuddin Omar. "Methods of Arabic language baseline detection–The state of art." IJCSNS 8.10 (2008): 137.

[37] Boukerma, Hanene, and Nadir Farah. "A novel Arabic baseline estimation algorithm based on sub-words treatment." Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on. IEEE, 2010.

[38] Osman, Yusra. "Segmentation algorithm for Arabic handwritten text based on contour analysis." Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on. IEEE, 2013.

[39] Al Hamad, Husam A., and Raed Abu Zitar. "Development of an efficient neural-based segmentation technique for Arabic handwriting recognition." Pattern Recognition 43.8 (2010): 2773-2798.

[40] Dinges, Laslo, Ayoub Al-Hamadi, and Moftah Elzobi. "A Locale Group Based Line Segmentation Approach for Non Uniform Skewed and Curved Arabic Handwritings." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[41] Eraqi, Hesham M., and Sherif Abdelazeem. "A new Efficient Graphemes Segmentation Technique for Offline Arabic Handwriting." Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on. IEEE, 2012.

[42] Lawgali, A., Bouridane, A., Angelova, M., & Ghassemlooy, Z. (2011, September). Automatic segmentation for Arabic characters in handwriting documents. In Image Processing (ICIP), 2011 18th IEEE International Conference on (pp. 3529-3532). IEEE.

[43] AlKhateeb, Jawad H., Jianmin Jiang, Jinchang Ren, and S. Ipson. "Component-based segmentation of words from handwritten Arabic text." International Journal of Computer Systems Science and Engineering 5, no. 1 (2009).

[44] Al Hamad, Husam A. "Neural-Based Segmentation Technique for Arabic Handwriting Scripts." 21st International Conference on Computer Graphics, Visualization and Computer Vision, WSCG (2013).

[45] Tamen, Zahia, and Habiba Drias. "How to overcome some segmentation problems in a constrained handwritten Arabic character recognition system." Information Sciences Signal Processing and their Applications (ISSPA), 2010 10th International Conference on. IEEE, 2010.

[46] Parvez, Mohammad Tanvir, and Sabri A. Mahmoud. "Lexicon Reduction Using Segment Descriptors for Arabic Handwriting Recognition." Document Analysis and Recognition (ICDAR), 2013 12th International Conference on. IEEE, 2013.

[47] Samoud, Fadoua Bouafif, Samia Snoussi Maddouri, and Hamid Amiri. "Three Evaluation Criteria's towards a Comparison of Two Characters Segmentation Methods for Handwritten Arabic Script." Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference on. IEEE, 2012.

[48] Haraty, Ramzi A., and Catherine Ghaddar. "Arabic text recognition." Int. Arab J. Inf. Technol. 1, no. 2 (2004): 156-163.

[49] Lawgali, A., & Bouridane, A. (2011). Handwritten Arabic Character Recognition: Which feature extraction method. International Journal of Advanced Science and Technology, 34(September)

[50] Naz, S., Hayat, K., Imran Razzak, M., Waqas Anwar, M., Madani, S. a., & Khan, S. U. (2014). The optical character recognition of Urdu-like cursive scripts. Pattern Recognition, 47(3), 1229–1248. doi:10.1016/j.patcog.2013.09.037

[51] Elbaati, Abdelkarim, Houcine Boubaker, Monji Kherallah, Abdel Ennaji, and A. M. Alimi. "Arabic handwriting recognition using restored stroke chronology." In *Document Analysis and Recognition, 2009. ICDAR'09. 10th International Conference on*, pp. 411-415. IEEE, 2009.

[52] Märgner, V., El, H., & Mario, A. Offline Handwritten Arabic Word Recognition Using HMM - a Character Based Approach without Explicit Segmentation. The institute for Communications Technology (IfN), Technical University of Braunschweig; Department of Signal Processing for Mobile Information Systems Schleinitzstrasse 22, 38106, Braunschweig, Germany.

[53] Al-Ohali, Yousef, Mohamed Cheriet, and Ching Suen. "Databases for recognition of handwritten Arabic checks." Pattern Recognition 36.1 (2003): 111-121.

[54] Al-Ma'adeed, Somaya, Dave Elliman, and Colin A. Higgins. "A data base for Arabic handwritten text recognition research." Frontiers in Handwriting Recognition, 2002. Proceedings. Eighth International Workshop on. IEEE, 2002.

[55] El Abed, Haikal, and V. Margner. "The IFN/ENIT-database-a tool to develop Arabic handwriting recognition systems." Signal Processing and Its Applications, 2007. ISSPA 2007. 9th International Symposium on. IEEE, 2007.

[56] 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

[57] 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. Märgner, V., & Abed, H. El. (2009). ICDAR 2009 Arabic Handwriting Recognition Competition. 2009 10th International Conference on Document Analysis and Recognition, 1383–1387. doi:10.1109/ICDAR.2009.256

[58] Margner, Volker, and Haikal El Abed. "ICFHR 2010-Arabic handwriting recognition competition." Frontiers in Handwriting Recognition (ICFHR), 2010 International Conference on. IEEE, 2010.

[59] Margner, V., & Abed, H. El. (2011). ICDAR 2011 - Arabic Handwriting Recognition Competition. 2011 International Conference on Document Analysis and Recognition, 1444–1448. doi:10.1109/ICDAR.2011.287

[60] 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

[61] Musa, M.E.M., "Arabic handwritten datasets for pattern recognition and machine learning," Application of Information and Communication Technologies (AICT), 2011 5th International Conference on , vol., no., pp.1,3, 12-14 Oct. 2011.

[62] Mahmoud, Sabri A., Irfan Ahmad, Mohammad Alshayeb, Wasfi G. Al-Khatib, Mohammad Tanvir Parvez, Gernot A. Fink, Volker Margner, and Haikal El Abed. "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.

[63] Mahmoud, Sabri A., Irfan Ahmad, Wasfi G. Al-Khatib, Mohammad Alshayeb, Mohammad Tanvir Parvez, Volker Märgner, and Gernot A. Fink. "KHATT: An open Arabic offline handwritten text database." Pattern Recognition 47, no. 3 (2014): 1096-1112.

[64] Lawgali, A., M. Angelova, and A. Bouridane. "HACDB: Handwritten Arabic characters database for automatic character recognition." *Visual Information Processing (EUVIP), 2013 4th European Workshop on*. IEEE, 2013.

[65] Dinges, Laslo, Ayoub Al-Hamadi, and Moftah Elzobi. "An Approach for Arabic Handwriting Synthesis Based on Active Shape Models." *Document Analysis and Recognition (ICDAR), 2013 12th International Conference on*. IEEE, 2013.

[66] Farah, Nadir, Labiba Souici, Lotfi Farah, and Mokhtar Sellami. "Arabic words recognition with classifiers combination: An application to literal amounts." In *Artificial Intelligence: Methodology, Systems, and Applications*, pp. 420-429. Springer Berlin Heidelberg, 2004.

[67] Farah, Nadir, Tarek Khadir, and Mokhtar Sellami. "Artificial neural network fusion: Application to Arabic words recognition." In *ESANN*, pp. 151-156. 2005.

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

[69] Souici-Meslati, Labiba, and Mokhtar Sellami. "A Hybrid Approach for Arabic Literal Amounts Recognition", Arabian Journal for Science & Engineering (Springer Science & Business Media BV) 29 (2004).

[70] Lotfi, Farah, Farah Nadir, and Bedda Mouldi. "Arabic Words Recognition by Fuzzy Classifier." *Journal of Applied Sciences* 6 (2006): 647-650.

[71] Amrouch, Siham, Aida Chefrour, and Labiba Souici-Meslati. "Decision Trees for Handwritten Arabic Words Recognition", Proceedings of the International Arab Conference on Information Technology (ACIT), Erriad-Saudi Arabia, 2011.

[72] Souici, Labiba, Nadir Farah, Toufik Sari, and Mokhtar Sellami. "Rule based neural networks construction for handwritten Arabic city-names recognition." In *Artificial Intelligence: Methodology, Systems, and Applications*, pp. 331-340. Springer Berlin Heidelberg, 2004.

[73] Rehman, Amjad, Dzulkifli Mohamad, and Ghazali Sulong. "Implicit vs explicit based script segmentation and recognition: a performance comparison on benchmark database." *Int. J. Open Problems Compt. Math* 2, no. 3 (2009): 352-364.

[74] Choudhary, Amit. "A Review of Various Character Segmentation Techniques for Cursive Handwritten Words Recognition." International Journal of Information & Computation Technology (IJICT), Volume 4, Number 6 spl. (2014)

[75] Koerich, Alessandro L., Robert Sabourin, and Ching Y. Suen. "Large vocabulary off-line handwriting recognition: A survey." *Pattern Analysis & Applications* 6, no. 2 (2003): 97-121.

[76] Kundu, Amlan, Tom Hines, Jon Phillips, Benjamin D. Huyck, and Linda C. Van Guilder. "Arabic handwriting recognition using variable duration HMM." In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2, pp. 644-648. IEEE, 2007.

[77] Al-Hajj, Ramy, Chafic Mokbel, and Laurence Likforman-Sulem. "Combination of HMM-based classifiers for the recognition of Arabic handwritten words." In *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, vol. 2, pp. 959-963. IEEE, 2007.

[78] Saleem, S., Cao, H., Subramanian, K., Kamali, M., Prasad, R., & Natarajan, P. (2009). Improvements in BBN's HMM-based offline Arabic handwriting recognition system. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR (pp. 773–777).

[79] Eraqi, H. M., & Abdelazeem, S. (2012). HMM-based Offline Arabic Handwriting Recognition: Using New Feature Extraction and Lexicon Ranking Techniques. 2012 International Conference on Frontiers in Handwriting Recognition, 554–559. doi:10.1109/ICFHR.2012.214

[80] Khorsheed, Mohammad S. "Recognizing handwritten Arabic manuscripts using a single hidden Markov model." *Pattern Recognition Letters* 24, no. 14 (2003): 2235-2242.

[81] Dehghan, Mehdi, Karim Faez, Majid Ahmadi, and Malayappan Shridhar. "Handwritten Farsi (Arabic) word recognition: a holistic approach using discrete HMM." *Pattern Recognition* 34, no. 5 (2001): 1057-1065.

[82] El-Hajj, Ramy, Laurence Likforman-Sulem, and Chafic Mokbel. "Arabic handwriting recognition using baseline dependant features and hidden markov modeling." In *Document Analysis and Recognition, 2005. Proceedings. Eighth International Conference on*, pp. 893-897. IEEE, 2005.

[83] Alma'adeed, Somaya, Colin Higgens, and Dave Elliman. "Recognition of off-line handwritten Arabic words using hidden Markov model approach." In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, vol. 3, pp. 481-484. IEEE, 2002.

[84] Elzobi, Moftah, Ayoub Al-Hamadi, Laslo Dings, Mahmoud Elmezain, and Anwar Saeed. "A Hidden Markov Model-Based Approach with an Adaptive Threshold Model for Off-Line Arabic Handwriting Recognition." In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on, pp. 945-949. IEEE, 2013.

[85] AlKhateeb, Jawad H., Jinchang Ren, Jianmin Jiang, and Husni Al-Muhtaseb. "Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking." Pattern Recognition Letters 32, no. 8 (2011): 1081-1088.

[86] Lajish, V. L. "A Quick Review of Recognition Strategies Based on Neural Network and Neuro-Fuzzy Approaches with Special Reference to HCR in Indian Languages." (2013).

[87] Asiri, A., and Mohammad S. Khorsheed. "Automatic Processing of Handwritten Arabic Forms using Neural Networks." In *IEC (Prague)*, pp. 313-317. 2005.

[88] Zaidan, A. A., B. B. Zaidan, Hamid Jalab, Hamdan Alanazi, and Rami Alnaqeib. "Offline Arabic Handwriting Recognition Using Artificial Neural Network." *arXiv preprint arXiv: 1006.2809* (2010).

[89] Alma'adeed, Somaya. "Recognition of off-line handwritten Arabic words using neural network." In Geometric Modeling and Imaging--New Trends, 2006, pp. 141-144. IEEE, 1993.

[90] Mohammed, Naji F., and Nazlia Omar. "Arabic named entity recognition using artificial neural network." *Journal of Computer Science* 8, no. 8 (2012): 1285.

[91] Perwej, Yusuf. "Recurrent Neural Network Method in Arabic Words Recognition System." *arXiv preprint arXiv: 1301.4662* (2013).

[92] Cheikh, I. Ben, Belaïd, A., Kacem, A., Esstt, U., Hussein, A. T., Mnara, B. P., & Nancy, V. L. (2008). A Novel Approach for the Recognition of a Wide Arabic Handwritten Word Lexicon Neural model : How to benefit from the, 2–5.

[93] Al Hamad, Husam Ahmed. "Use an efficient neural network to improve the Arabic handwriting recognition." In *Signal and Image Processing Applications (ICSIPA), 2013 IEEE International Conference on*, pp. 269-274. IEEE, 2013.

[94] Zadeh L. A., "Fuzzy sets," Information and Control, Vol. 8, pp. 338-353, 1965.

[95] Zadeh L. A., "Fuzzy Algorithms," Information and Control, Vol. 12, pp. 94-102, 1968.

[96] Abuhaiba, Ibrahim SI, Sabri A. Mahmoud, and Roger J. Green. "Recognition of handwritten cursive Arabic characters." Pattern Analysis and Machine Intelligence, IEEE Transactions on 16.6 (1994): 664-672.

[97] Baccour, Leila, and Adel M. Alimi. "A comparison of some intuitionistic fuzzy similarity measures applied to handwritten Arabic sentences recognition." Fuzzy Systems, 2009. FUZZ-IEEE 2009. IEEE International Conference on. IEEE, 2009.

[98] Leila, Chergui, Kef Maamar, and Chikhi Salim. "Combining neural networks for Arabic handwriting recognition." Programming and Systems (ISPS), 2011 10th International Symposium on. IEEE, 2011.

[99] Nemouchi, Soulef, Labiba Souici Meslati, and Nadir Farah. "Classifiers combination for Arabic words recognition: application to handwritten Algerian city names." Image and Signal Processing. Springer Berlin Heidelberg, 2012. 562-570.

[100] Dehghan, Mehdi, Karim Faez, Majid Ahmadi, and Malayappan Shridhar. "Unconstrained Farsi handwritten word recognition using fuzzy vector quantization and hidden Markov models." Pattern Recognition Letters 22, no. 2 (2001): 209-214.

[101] Pach, Ferenc Peter, Janos Abonyi, Sandor Nemeth, and Peter Arva. "Supervised clustering and fuzzy decision tree induction for the identification of compact classifiers." In *5th International Symposium of Hungarian Researchers on Computational Intelligence, Budapest, Hungary*. 2004.

[102] MATIAŠKO, Karol, Ján BOHÁČIK, Vitaly LEVASHENKO, and Štefan KOVALÍK. "Learning fuzzy rules from fuzzy decision trees." *Journal of Information, Control and Management Systems* 4, no. 2 (2006).

[103] Mitra, S., Member, S., Konwar, K. M., & Pal, S. K. (2002). Fuzzy Decision Tree, Linguistic Rules and Fuzzy Knowledge-Based Network : Generation and Evaluation, *32*(4), 328–339.

[104] A. Amin, "Recognition of printed Arabic text based on global features and decision tree learning techniques", Pattern recognition, Vol. 33, No. 8, pp. 1309-1323, 2000.

[105] I. S. I. Abuhaiba, "Arabic font recognition using decision trees built from Common Words", Journal of Computing and Information Technology: CIT, Vol. 13, Num. 3, pp. 211-223, 2005

[106] D.S. Guru, S.K. Ahmed, K. Irfan, An attempt towards recognition of hand-written Urdu characters: a decision tree approach, in: Proceedings of the National Conference on Computers and Information Technology (NCCIT'01), 2001, pp. 75–83.

[107] Chowdhury Mofizur Rahman and Monzur Morshed " Decision tree based learning of handwritten Bangla characters, ICCIT 99, SUST, Sylhet, December 3- 5,1999.

[108] Chang, Robin L P; Pavlidis, Theodosios, "Fuzzy Decision Tree Algorithms," Systems, Man and Cybernetics, IEEE Transactions on , vol.7, no.1, pp.28,35, Jan. 1977 doi: 10.1109/TSMC.1977.4309586 decision tree

[109] Fang, G., Gao, W., & Zhao, D. (2004). Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 34(3), 305–314. doi:10.1109/TSMCA.2004.824852

[110] Rangkuti, A. Haris, Nashrul Hakiem, Rizal Broer Bahaweres, Agus Harjoko, and Agfianto Eko Putro. "Analysis of image similarity with CBIR concept using wavelet transform and threshold algorithm." In *Computers & Informatics (ISCI), 2013 IEEE Symposium on*, pp. 122-127. IEEE, 2013.

[111] Yang, Shiueng-Bien, and Tzu-Wei Chen. "Fuzzy variable-branch decision tree for speech recognition." In *Asian Control Conference, 2009. ASCC 2009. 7th*, pp. 773-778. IEEE, 2009.

[112] Tabakov, Martin, M. Podhorska-Okolow, S. Zareba, and Bartosz Pula. "Using fuzzy Sugeno integral as an aggregation operator of ensemble of fuzzy decision trees in the recognition of HER2 breast cancer histopathology images." In *Computer Medical Applications (ICCMA), 2013 International Conference on*, pp. 1-6. IEEE, 2013.

