Volume 4, Issue 10, October 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Isolated Arabic Handwritten Character Recognition: A Survey

Omar Balola Ali
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, USA

Abstract- Offline Arabic handwriting character recognition (AHCR) systems are very important since they make life easier for governments, researchers and scholars who are dealing with Arabic language in education, documentation and security. A widening use of the Arabic script in countries that deals with the Arabic language and countries that use the Arabic script in their languages such as Persian and Urdu makes offline Arabic handwriting character recognition a necessary system to have. Some of the benefits of such a system would be in processing checks, converting handwritten text into printed text, processing handwritten reports etc. The need for offline AHCR systems are more nowadays because of the expansion of technology and the convenience for customers. Many AHCR algorithms have been designed and implemented using various types of technologies which helped in reaching high recognition rate of accuracy. This paper presents a survey of the research published in this area. The paper will analysis and compare the various algorithms with respect to different stages of the offline AHCR. Preprocessing methods, feature extraction techniques and different classification approaches will also be presented. Future research in Arabic handwriting recognition will be discussed and analyzed. The paper also presents a new proposed two stages neuro-fuzzy approach for isolated Arabic handwritten character recognition system.

Keywords- Offline Arabic Character Recognition, Genetic Algorithms, Fuzzy Logic, Neural Network, Fuzzy-Neural Systems

I. INTRODUCTION

Handwritten Arabic Character Recognition system (HACR) is the system that attempts to recognize a text that has been written by a person (not a machine). Handwriting recognition attracted the attention of researchers since the origin of computers. Nowadays, the technological progress made in the field of computer architectures and peripheral devices as well as the advances in scientific research make it possible for the development of new systems for handwriting recognition. In addition, the series of Workshops on Frontiers in Handwriting Recognition and the International Conferences on Document Analysis and Recognition provide continuously new stimuli to researchers which points out that handwriting recognition industry can become a large business [1]. The goal of a character recognition system is to transform a character handwritten on paper into a digital format that can be manipulated by characters processor software.

The handwriting recognition (HWR) systems can be classified into two main groups, off-line and on-line recognition, according to the format of handwriting inputs. In offline recognition, only the image of the handwriting is available, while in the on-line case temporal information such as pen tip coordinates, as a function of time, is also available. Many applications require off-line HWR capabilities such as bank processing, mail sorting, document archiving, commercial form-reading, office automation, etc. [2].

Fuzzy logic was first introduced by Lotfi Zadeh [3]. It was developed for solving decision making problems through the use of "IF-THEN" rules. It was used later to model uncertainty and imprecision in data management. Fuzzy logic is an easy way to reach definitive conclusions based on vague, ambiguous, imprecise and noise information [4] [5].

Neural Network approach is an emerging technique in the area of handwritten character recognition through the use of Artificial Neural Network (ANN) implementations were networks employs specific learning rules to update the links (weights) among their nodes. Such networks can be fed with data from an input picture and trained to output characters in one or more forms [6]. There are many structures of ANNs including, Perceptron, Adaline, Madaline, Kohonen, Back Propagation and many others. Back Propagation ANN is the most commonly used since it is effective and very simple to implement [7].

Hybrid approaches could be considered as one of the main contributions of soft computing with neuro-fuzzy systems being the first and probably the most successful hybrid approach till now. Neuro-fuzzy systems incorporate the elements from fuzzy logic (FL) and neural networks (NN). The idea of hybridization originates from the following two observations:

- Fuzzy systems are not capable of parallel computation, whereas these characteristics are clearly attributed to NNs.
- NNs lack flexibility, human interaction which lies at the core of FL.

In this paper we will be proposing a handwritten recognition system for Arabic characters using an artificial neural network fuzzy inference system (ANFIS).

The paper is organized as follows: section 2 presents a brief introduction to the characteristics of Arabic writing, section 3 presents a literature review for Arabic Handwritten Character Recognition Systems, section 4 presents an overview of existing neuro-fuzzy algorithms for Arabic handwriting, section 5 will introduce a new isolated Arabic handwritten character recognition system and the conclusion is presented in section 6.

II. CHARACTERISTICS OF ARABIC WRITING

A. Handwritten Character Recognition Position

The Arabic alphabet contains of 28 letters. Each has between two and four shapes, depends on the position of the letter within the word or sub-word. The shape corresponds to one of four position; beginning, middle, end and isolated as shown in table 1. Beginning is where the character is positioned at the beginning of the word (the first letter of the word). Middle is where the character is located between two characters in a word. End is where the character is positioned at the end of the word. Isolated is where the character is located disconnected at the beginning of the word or at the middle of the word or at the end of the word like the characterأ.

B. Arabic Handwritten Character Difficulties

The following are difficulties with the Arabic handwritten recognition:

- Arabic characters can have more than one shape according to their position in a word whether at the beginning, middle, end, or stand alone.
- Different writers or even the same writer under different conditions will write some Arabic characters in completely different ways.
- Other characters have very similar contours and are difficult to recognize especially when non-character and external objects are present in the scanned image.

Figure 1 shows a list of the Arabic characters. Handwritten Arabic characters depend largely on contextual information.

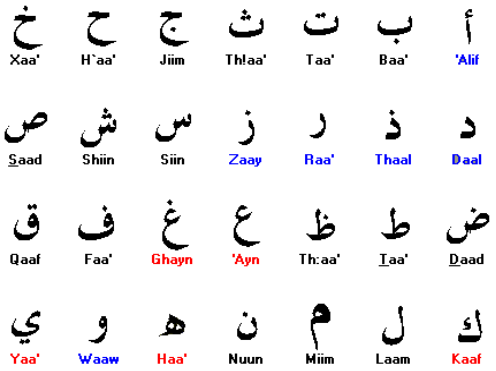

Figure 1. Arabic characters.

Table 1. Arabic characters based on their position within a word.

	Character	Isolated	Begin	Middle	End
Alef	الف	ا	ا	ـا	ـا
Ba'	باء	ب	بـ	ـبـ	ـب
Ta'	تاء	ت	تـ	ـتـ	ـت
Tha'	ثاء	ث	ثـ	ـثـ	ـث
Jeem	جيم	ج	جـ	ـجـ	ـج
H'a'	حاء	ح	حـ	ـحـ	ـح
Kha'	خاء	خ	خـ	ـخـ	ـخ
Dal	دال	د	د	ـد	ـد
Thal	ذال	ذ	ذ	ـذ	ـذ

Raa'	راء	ر	ر	ـر	ر
Zain	زين	ز	ز	ـز	ز
Seen	سين	س	سـ	ـسـ	س
Sheen	شين	ش	شـ	ـشـ	ش
Sad	صاد	ص	صـ	ـصـ	ص
Dad	ضاد	ض	ضـ	ـضـ	ض
Taa	طاء	ط	طـ	ـطـ	ط
Zain	ظين	ظ	ظـ	ـظـ	ظ
Ain	عين	ع	عـ	ـعـ	ع
Gain	غين	غ	غـ	ـغـ	غ
Fa	فاء	ف	فـ	ـفـ	ف
Gaf	قاف	ق	قـ	ـقـ	ق
Kaf	كاف	ك	كـ	ـكـ	ك
Lam	لام	ل	لـ	ـلـ	ل
Mim	ميم	م	مـ	ـمـ	م
Non	نون	ن	نـ	ـنـ	ن
Ha	هاء	هـ	هـ	ـهـ	ه
Waw	واو	و	و	ـو	و
ya	ياء	ي	يـ	ـيـ	ي

III. LITERATURE REVIEW- ARABIC HANDWRITTEN CHARACTER RECOGNITION SYSTEMS

Character recognition goes through several stages as shown in figure 2. The first stage is a scanned text page. There are some preprocessing steps followed after the scanned page, where attempt to improve the performance of letter recognition. The preprocessing techniques, which are resized, noise removal, median filter and grayscale, skew detection and correction, etc. After preprocessing, the text image will be segmented into many shapes, lines, words, sub-words, characters or sub characters. The next stage is to extract features. Features are usually used as inputs to train the classifier to build the models (training and testing) through the classification stage. The last stage in the recognition system is the post-processing stage. Post-processing improves recognition.

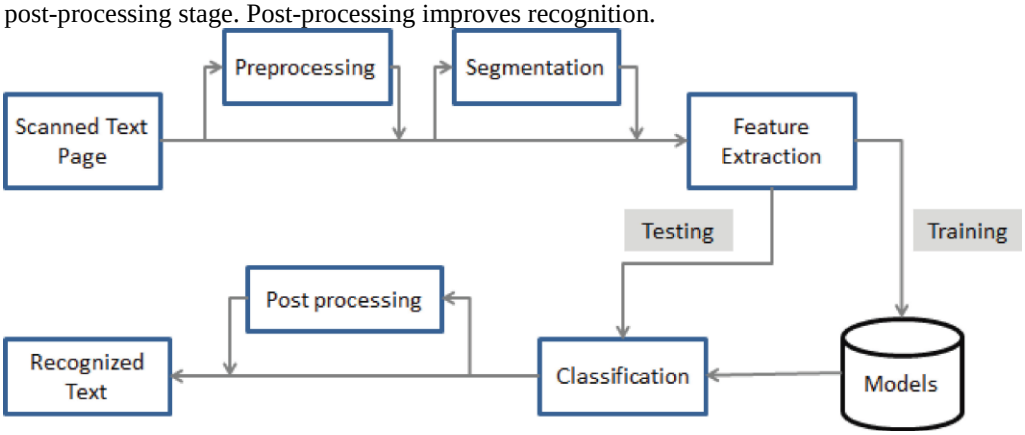

Figure 2. Character recognition stages

A. Preprocessing

Preprocessing stage involves series of techniques performed to enhance the image to make it suitable for segmentation. The Preprocessing techniques, noise removal, normalization, skew correction, thinning and slant. The aim of preprocessing stage is the removal of all elements in the word image that are not useful in the recognition process.
The research that is related to the preprocessing stage for Arabic handwritten text images can be classified as follows:

- Thresholding- This technique involves transforming an input image from a gray image into a binary image and was cited by [11, 12, 21, 23, 23, 30 and 37] to get a good features from the image,
- Thinning - The thinning technique will transform the binary image of the character to a one pixel thick image. Hanan et al [9], Albakooret al [26] and Namazi et al have used this technique in their researches,
- Normalization - Normalization is the process to converting the random size for images into standard size. Found in the research [12] to a common defined size, example 30x30 pixels matrix,
- Noise Removal - The major objective of noise removal is to remove any unwanted bit-patterns which do not have any significance in the output. The Median filtering technique was use by [10,12,15,19 and 31] to remove noise,
- Skew detection and skew correction technique used by [10 and 15],
- Skeletonization used by [10 and 26],
- Smoothing used by [13,23, 28 and 30],
- Filterization used by [12 and 21], and
- Horizontal projection profile used by [9 and 16].

B. Feature Extraction

Feature extraction is the heart of any pattern recognition application. The traditional goal of the feature extractor stage is to characterize an object by making numerical measurements. Good features are those whose values are similar for objects belonging to the same category and distinct for objects in different categories. Feature extraction is very problem dependent [38]. Features are will extracted from the image of a word/character which is expected to represent the shape of the images. There are a number of famous features that are related to the character recognition problem. There are numerous types of features proposed by researchers. There are two types of features; Structural features such as loops, branch points, endpoints, dots, etc. and Statistical features such as pixel densities, histograms of chain code directions, moments and Fourier descriptors.

1) Structural features

These types of features were used by [9, 18, 24, 27 and 34]. In [9] structural features that were used are the following: the length of the character, the width of it, if it has a loop or not, if there is a right character to connect to it, if there is a left character to connect to and if there is a complement of the character like the zigzag shape (Hamza), one point, two point or three point. In [18] they used features such as loops, holes, strokes, vertical lines and cusps.

2) Statistical features

Pixel densities features were use in [8, 14 and 17]. Sum of black and white pixels, number of pixels were used in [14] and in [17] detecting black and white points were used as statistical features. Moment invariants were used by [10 and 29]. In [10] they proposed seven invariant moments; example of these moments is invariant under reflection, rotation and scaling. In [29] they used scale, orientation and center of gravity. Vertical and horizontal projections were used in [9 and 16]. In [9] the longest spike which represented the baseline was used. In [26] branch, start-point and end-point of a character features were used. Wavelet transforms features were used by [31 and 32]. In [11 and 20] they used the following features: determine the body and secondary part, position of the part above or below, loop and Radon transform of the characters were used. Code of chain feature used in [25] and Pseudo-Zernike Moments, size, rotation and translation invariant features were also used in [37]. Table 2 summarizes some of the preprocessing techniques and features extraction methods used in offline Arabic text recognition.

C. Classification Approaches

Classification is the main stage of AHCR system. It uses the features in the feature extraction stage as inputs to the model to identify the text segment according to stated rules. These include Hidden Markov Model (HMM), Support Vector Machines (SVM), Artificial Neural Networks (ANN), k-Nearest Neighbors (k-NN), fuzzy logic, genetic algorithm and others. The different approaches for classify character recognition will presented in the following section.

Table 2. Summary of some of the preprocessing techniques and features extraction methods, used in offline Arabic text recognition.

Preprocessing	Features	Reference, date
	Sum of black and white pixels.	[8], 2010
Thinning technique for Arabic words and finds the vertical and horizontal projection profile.	the length of the character, the width of it, if it has a loop or not, if there is a right character to connect to it, if there is a left character to connect to and if there is a complement of the character like the zigzag shape (Hamza), one point, two points or three points.	[9], 2010

Noise analysis and removal, Skew detection and correction and Evaluating the gap between the words and characters.	The features in each window are extracted through seven invariant moments example of these moments are invariant under reflection, rotation and scaling.	[10], 2011
Binary, filtering, and skeleton.	divide the letters into regions for determine the body and secondary part, position of the part above or below, loop and Radon transform.	[11], 2011
Filter, binary image and resized.	Width of the dynamic-sized window and the height.	[12], 2012
Smoothing and de-noising.	Different the values, similar value and Independence features.	[13], 2009
	Width, length, number of pixels and height-to-width ratio.	[14], 2005
Removal of isolated pixels, skew detection and correction.	Invariant under image translation, scaling and rotation.	[15], 2003
Vertical and horizontal projection.	Horizontal and vertical projection profile and Laplacian filter.	[16], 2012
Convert RGB image and crop the white area around the word.	Detecting black and white points.	[17], 2012
Black and white format and noise.	Loops, holes, strokes, vertical lines, and cusps.	[18], 2008
	Genetic algorithms representing a subset of evolutionary computation techniques.	[19], 2008
	Secondary components, Main body, Skeleton and Boundary feature.	[20], 2008
Resizing, filtration, and converting to binary.	Ratio of white pixels to black pixels, ratio between the two farthest pixels, average of the spatial segment, variance of the farthest two vertical pixels, average value of the line and total variance of the segment pixels.	[21], 2012
	Aspect and stroke ratio.	[22], 2009
Binarization and smoothing.	Discrimination, reliability and independence	[23], 2001
Binary images.	The contours (dots and holes).	[24], 2006
	Morphological primitives and the statistic primitives.	[25], 2006
Skeletonization and thinning	Branch and start-point and end-point of a character or an end-point and start-point of a straight piece in a character.	[26], 2009
	Number of dots in the character and where the dot is, if it's up, down or in the middle of the character.	[27], 2010
Smoothing operation and cursive characters.	The velocity profile of each stroke.	[28], 1997
	Moment invariants, scale, orientation and center of gravity	[29], 2007
Binary and smoothing.	Uniform and no-uniform.	[30],2006
Noise reduction and edges detection.	Hough transform and wavelets transform.	[31], 2006
	Statistical features from the standard deviation of the wavelet coefficients besides the density of black pixels.	[32], 2006
	PCA for feature extraction, calculate the Eigenvectors and Eigen values.	[33], 2007
	Structural features and Direction feature.	[34], 2013
	- Dimensions of the including rectangle (high and width).	[35], 2009

	- Relative position of the including rectangle - Positions of the references points. - Direction of the trajectory on the level of the point's references.	
	Velocity profile of each stroke.	[36], 1997
Binarization and thinning.	Pseudo-Zernike Moments, size, rotation and translation invariant.	[37],1996

1) Hidden Markov Models (HMM)

Many researchers have used (HMM) for Arabic text recognition [30, 31 and 35]. The general trend of using HMM is to use a sliding window of the text line image to convert a 2-dimensional image to a 1-dimensional feature vector [39]. In [30] they used a HMMs algorithm with different explicit distribution for the state duration and after combining statistical and structural features the best recognition rate was 91.23%. In [31] they used wavelet transform features extraction and presented the edge detection of the character features with Hough Transform. In [35] they used genetic algorithm (GA) to optimize the sequences of handwritten strokes with HMM classifier.

2) Fuzzy logic approach

Fuzzy logic is a soft computing approach based on "degrees of truth" rather than the usual "true or false" logic. Many researchers have applied FL for recognition [8, 14, 25 and 29].

3) Neural network

The main driving force behind neural network research is the desire to create a machine that works similar to the manner in which our own brain works. Neural networks have been used in a variety of different areas to solve a wide range of problems. Unlike human brains that can identify and memorize the characters like letters or digits; computers treat them as binary graphics. Therefore, algorithms are necessary to identify and recognize each character [40].

Neural networks were used for Arabic text recognition [13, 19, 21, 26 and 34]. Back-propagation algorithm was used in [13] for recognition. In [19] they proposed two neural networks, a Multi-Layer Perceptron (MLP) and a Learning Vector Quantization (LVQ). In [21] they present cascaded networks to recognize the characters with 120 features for each character's image. Back propagation Neural Network algorithm was also used in [26] with histogram features. In [34] authors have designed Radial-Basis neural network classifier, investigated and compared among results of four different artificial neural network models.

4) K-Nearest Neighbor

K-nearest neighbor algorithm (k-NN) [41 and 42] is a method for classifying objects based on closest training examples in the feature space. The K-NN algorithm is amongst the simplest of all machine learning algorithms; an object is classified by a majority vote of its neighbors with the object being assigned to the class most common amongst its k nearest neighbors (k is a positive integer, typically small integer).

5) Neuro-Fuzzy

We found many researchers that have used Neuro-Fuzzy approach [18, 28, 32, 36 and 37]. In [11] they used Fuzzy logic with ART neural networks.

Table 3 shows a summary of some of the comparison of offline Arabic character recognition algorithms.

Table 3.Summarizes some of the recognition algorithm with respect to the dataset and accuracy in offline Arabic text recognition.

Algorithm Type	Data set	Accuracy	Reference, date
Fuzzy logic	96 different shapes of characters.	88%	[8], 2010
Genetic algorithm	15000 handwritten words written by 100 writers.	87%	[9], 2010
Minimum distances and cosine (Ø)	Collected by forty subjects.	83.09%, 92.86 %.	[10], 2011
Multilevel classifier		{93, 84, 89 and 85%} Isolated, beginning, middle and end.	[11], 2011
Dynamic-size windowing	100 character shapes, from many articles in Alrai newspaper.	96%	[12], 2012
Back-propagation	100 letters	95%	[13], 2009
Fuzzy logic			[14], 2005
Radial-Basis Function	149 characters	73%	[15], 2003

RBF Network			
Production rule, NN and decision tree	504 characters for training, 336 for testing	97.6%, 97.5% and 98.8%	[16], 2012
Matched data with black and white pixels	IFN/ENIT	81%	[17], 2012
Fuzzy ART neural network	700 sample characters	95%	[18], 2008
Multi-Layer Perceptron (MLP), Learning Vector Quantization (LVQ) and K–Nearest Neighbor (KNN)	2000 instances	98%, 96.54 and 97.15	[19], 2008
Linear Discriminant Analysis (LDA)	104 characters	87%	[20], 2008
Cascaded neural networks	100 different separated characters	68.10%	[21], 2012
Semi-supervised	IFN/ENIT	67% and 77%	[22], 2009
based segmentation technique		90%	[23], 2001
Fourier Spectrum (MFS)	78 Arabic characters	95.6%	[24], 2006
Fuzzy Logic (FL) and Expert System (ES)	280 characters	98.97%	[25], 2006
Back propagation Neural network		98.7%	[26], 2009
Decision tree	336 samples	97.6%	[27], 2010
Fuzzy neural network	more than 3000 characters	89%	[281], 1997
Particle swarm optimization	448 samples	82%	[29], 2007
a HMMs algorithm	26459 Arabic words written by 411 different writers	91.23 %.	[30],2006
Hidden Markov Models	170.000	96%	[31], 2006
Neuro-fuzzy	174394 Arabic characters in 9 different fonts.	95.64%.	[32], 2006
Nearest Neighbor	200 samples	90%.	[33], 2007
Radial-Basis neural	620 samples	95.32%	[34], 2013
Hidden Markov Model (HMM)		81% and 83%	[35], 2009
Fuzzy neural network	2000 characters written by the one writer	89%	[36], 1997
Neural networks and Fuzzy neural networks	3700 character samples.	99.85%.	[37],1996

IV. NEURO-FUZZY ALGORITHMS AND HANDWRITING

The integration of neural networks and fuzzy logic inference systems could be formulated into three main categories; cooperative, concurrent and integrated neuro-fuzzy models. In this section we will present the neuro- fuzzy algorithms used for handwriting recognition systems.

In 2010, Majidaet al [8] proposed a system to recognize isolated Arabic characters using fuzzy logic approach. The stages of classification that they used were three. The first is sum of pixels stage in which they determine the sum of black and white pixels and saved these matrices in different files. The second stage is templates comparison in which matching between rows is performed by comparing every template against every other saved template. In the last stage they used the matrices as input to the fuzzy system [8].

In 2009, Albakoor et al [26] proposed two important concepts; segmentation on the basis of word histogram and baseline estimation. Preprocessing stage techniques that were used are Skeletonization and thinning. The segmentation technique that was used here depended on the histogram observations of lines and columns included in the histogram of the word.

The words are segmented into their characters. Feature extracted with the following characteristic points: branch, start-point and end-point of a character or an end-point and start-point of a straight piece in a character. In classification they used Back propagation Neural Network algorithm, the recognition rate was 98.7% [26].

In 1997, Adel [28] described an on-line Arabic handwritten characters system. In this system, he used a fuzzy neural network for classifying characters. Preprocessing methods used were smoothing operation and cursive characters. The features extracted from each character are the neuro physiological parameters of the equation describing the curvilinear velocity of the script. The recognition stage is the Beta fuzzy neural network which has 100 neurons in the hidden layer and 55 neurons in output layer. The hierarchical neural network structure that he used contained six stages of three layers each. The system was implemented in C/C++ language and achieved a recognition rate of 89%.

In 2006, [32] proposed an Arabic characters recognition system using wavelet transform to extract features. Authors have used a neuro-fuzzy approach for character recognition. They used Fuzzy logic as a tool for enhancing the ability to deal with the recognition problem. The purpose of fuzzy features was to map the extracted features to values from 0 to l using a set of input membership functions. They used the Mamdani inference system. They tested two models of neural network, multilayer perceptron's (MLP) and radial basis function (RBF) networks. They proposed the Takagi-Sugeno integrated neuro-fuzzy system. The neural network learning algorithms was used to determine the parameters of the fuzzy inference system. The recognition rate was 95.64%.

In 1997, they proposed an on-line Arabic handwriting character system using fuzzy neural network as classifier. The system used a genetic algorithm to select the best combination of characters. Each character was represented by 6 feature vectors of n elements each (the n parameters Pi of the velocity profile of each stroke). The data set used for training was 2000 characters written by the one single writer [36].

In 1996, they designed a recognition system for recognizing printed Farsi/ Arabic characters with various fonts. They used Pseudo-Zernike Moments as input features which has been used for size, rotation and translation invariant. In the recognition stage, they used a combination of neural networks and fuzzy neural networks and this stage was done in two phases. In the first phase the output of these two networks were compared. In the second phase the results were corrected using 3700 character as training dataset. The Neural network had 36 inputs, 40 hidden nodes and 18 output nodes samples. Total accuracy was 99.35% [37].

V. A NEW PROPOSED APPROACH

In our approach, we will use two phases to deal with the issue of isolated Arabic handwritten character recognition; through combining two levels based on tow classifier, the public and the private according to the similar features between characters.

First, determine the similar of the Arabic letters. We will conduct recognition experiments of Arabic handwritten letters dataset. Each letter represents a different class, with neural fuzzy classifier. After the recognition result, conducting pro-processing work with confusion matrix, this determines the similar between the letters. Then will collect the similar letters in one class as shown in figure 3(a), with equal numbers for each letter. After that we will design classifiers for each class of the similar letters as shown in figure 3(b). When you run a full system first determines the character belonging to any separation of the similar letters and then in the second stage is to determine the specific letter. All classifiers will design with neuro-fuzzy algorithm as shown in figure 4.

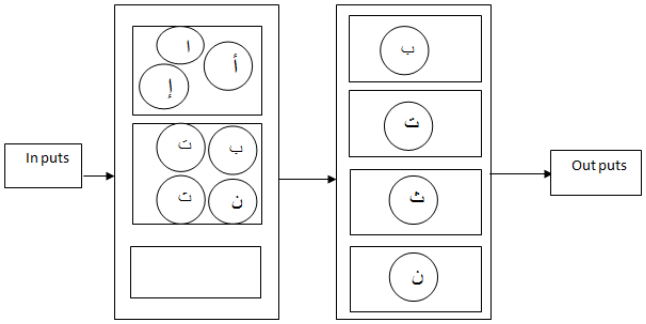

Figure 3(a) Figure 3(b)

Figure 3 (a) illustrate dataset of all Arabic character, each class concluded the similar characters. Figure 3 (b) illustrates dataset of the similar Arabic character, each class concluded one character.

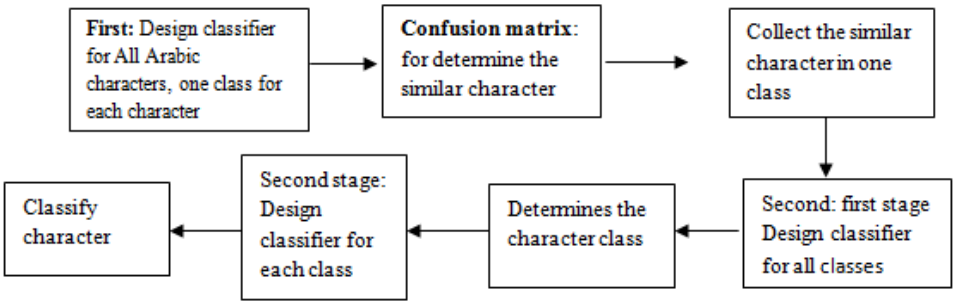

Figure 4. A new proposed approach with two stages of classifiers

In our proposed system for two stages-classifier system classification of Arabic handwritten isolated character, research methodology depends on four main stages are included, preprocessing, feature extraction, classification and pro-processing as shown in figure 5. The preprocessing is utilized to remove noise and after that extracts proper image features from each character. Then these features are used for classification to identify the character, neuro-fuzzy algorithm which will apply in two stages of classifiers. In the last we examine the result of testing dataset with confusion matrix.

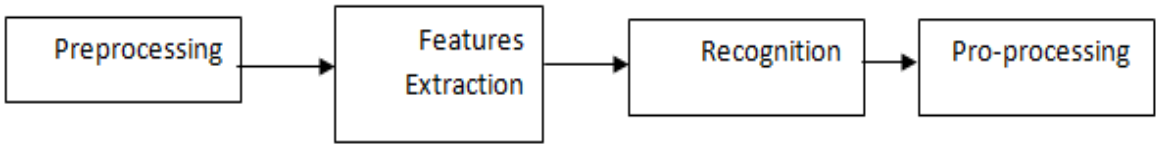

Figure 5. Illustrates the AHCR system.

In our research will use SUST-ARG dataset. SUST-ARG stands Sudan University for Sciences and Technology Arabic Recognition Group (SUST-ARG). The preprocessing techniques which will use are resized, noise removal and grayscale. The feature extraction methods that we will use are statistical features from the Standard deviation of the wavelet coefficients besides the density of black pixels. We will use the integration of neural networks and fuzzy systems. Some neural network models such as the back-propagation have been successfully applied to the training of neuro-fuzzy networks [32]. We will use the back-propagation neural network algorithm with six layers. We will define only three classes of objects with membership function. These classes are Negative zero (N), Zero (Z) means moments between -1 and 1 and Positive (P): moments equal to or greater than zero [37].

VI. CONCLUSION

The goal of this paper was to provide a detailed survey of published research work in the sequential stages of Arabic handwriting character recognition. We have found that various characteristics such as feature extraction, preprocessing, data set, segmentation and recognition rate are used to characterize the of offline Arabic handwriting text recognition systems.

Table 2 summarizes some of the preprocessing techniques and features extraction methods used in offline Arabic text recognition. Table 3 summarizes some of the recognition algorithm with respect to the dataset and accuracy in offline Arabic text recognition. There are different techniques for preprocessing such as Thinning, Binarization, noise removal, and Smoothing operation.

There are different approaches for feature extraction such as Principal Component Analysis, Horizontal, vertical projection, invariant moments and Hough transform and wavelets transform.

As can be seen from Table 3 many of the recent results for Arabic text recognition algorithms are reported with different training dataset and scope of recognition. Although researchers have designed different techniques for Arabic text recognition, these different techniques have used different databases for Arabic text recognition to evaluate their performance.

In Arabic text recognition approaches there are many developed algorithms for isolated character recognition systems such as neural networks (NNs), hidden Markov models (HMM), fuzzy set and neuro-fuzzy. We also found many systems were designed using multilevel classifiers with combination of the neural network and fuzzy logic.

There are many similarities between Arabic characters in terms of structural and morphology. There are many number and position of dots that differentiate between the otherwise similar characters like (ح , ج and خ) and (ب, ن, ت). This is the similarity between the letters that makes the Arabic character recognition difficult. We are proposing an algorithm which will combine two or more techniques to improve the accuracy of the AHCR system and we propose to use many levels of classifying stages to improve the recognition rate.

REFERENCES

[1] S. Impedovo, "Frontiers in Handwriting Recognition", in "Fundamentals in Handwriting Recognition", S. Impedovo (ed.), NATO-ASI Series, Springer- Verlag Publ., Berlin, 1994, G. Dimauro, S. Impedovo, G. Pirlo, A. Salzo Handwriting Recognition: State of the Art and Future Trends.

[2] H. A. Aljuaid and D. Muhamad, "Offline Arabic Character Recognition using Genetic Approach: A Survey".

[3] M. A. Abed, H. A. Abed, "Fuzzy Logic approach to Recognition of Isolated Arabic Characters," International Journal of Computer Theory and Engineering, Vol. 2, No. 1 February, 2010 1793-8201pp.9-17.

[4] R. E, "Introduction to Fuzzy set theory basic concepts and structures", IEEE, 1992.

[5] Z. Lotfi, "Fuzzy logics systems: origin concepts, and trends", 2004.

[6] "Neural Networks for Unicode Optical Character Recognition", www.projectsatbangalore.com/IEEE.pdf.

[7] "α-Soft: An English Language OCR" (Computer Engineering and Technology (ICCET), 2nd International Conference on 16-18 April 2010), 2010.

[8] M .A. Abed, H. A. Abed AL, Zainab Baha Al-Deen, Ahmad N. Ismail, "Fuzzy Logic approach to Recognition of Isolated Arabic Characters", International Journal of Computer Theory and Engineering, Vol. 2, No. 1 -1793 8201, February, 2010.

[9] H. Aljuaid, Z. Muhammad and M. Sarfraz, "A Tool to Develop Arabic Handwriting Recognition System Using Genetic Approach", Journal of Computer Science 6 (6): 597- 602ISSN 1549-3636, 2010.

[10] M. O. Al_Barraq and S.C.Mehrotra, "Handwritten Arabic Text Recognition System using Window Based Moment Invariant Method", international journal of advanced research in computer science, Vol. 2, Number 1, jan-feb 2011.

[11] R. I. Zaghloul, E. F. Alrawashdeh, D. Mohammad, and K. Bader, "Multilevel Classifier in Recognition of Handwritten Arabic Characters" Journal of Computer Science 7 (4): 512-518, ISSN 1549-3636, 2011.

[12] A. Mesleh1, A. Sharadqh, J. Al-Azzeh, M. Abu-Zaher, N. Al-Zabin, T. Jaber, A. Odeh and M. Hasn, "An Optical Character Recognition", Contemporary Engineering Sciences, Vol. 5, no. 11, 521 – 529, 2012.

[13] X. Wang, T.Huang and X. Liu, "Handwritten Character Recognition Based on BP Neural Network", Third International Conference on Genetic and Evolutionary Computing, 2009.

[14] M. Z. Khedher. G. Al-Talib," Recognition of secondary characters in handwritten Arabic using Fuzzy Logic ", International Conference on Machine Intelligence (ICMI'05), Tozeur, Tunisia, 2005.

[15] M. Sarfraz, S. N. Nawaz and A. Al-Khuraidly," Offline Arabic Text Recognition system" , IEEE International Conference on Geometric Modeling and Graphics (GMAG'03) 0-7695-1985-7/03, 2003.

[16] S. M. Ismail, S. Norul, and H. Sheikh," Online Arabic Handwritten Character Recognition Based On A rule Based Approach" ,Journal of Computer Science, 1859- 1868 ISSN 1549-3636, 8 September 2012.

[17] M. A. Abdullah, L. M. Al-Harigy, and H. H. Al-Fraidi," Off-Line Arabic Handwriting Character Recognition Using Word Segmentation ", JOURNAL OF COMPUTING, Vol. 4, ISSUE 3, ISSN 2151-9617, MARCH 2012.

[18] H. Almohri, J. S. Gray and H. Alnajjar, PhD, "A Real-time DSP-Based Optical Character Recognition System for Isolated Arabic characters using the TI TMS320C6416T", The IAJC-IJME International Conference ISBN 978-1-60643-379-9, 2008.

[19] L. Cordella C. D Stefano, F. Fontanella and C. Marrocco, "A Feature Selection Algorithm for Handwritten Character Recognition", IEEE 978-1-4244-2175-6/08, 2008.

[20] G. A. Abandah, K. S. Younis and M. Z. Khedher, "Handwritten Arabic Character Recognition Using Multiple Classifiers Based ON Letter Form" In Proc. 5th IASTED Int'l Conf. on Signal Processing, Pattern Recognition, & Applications (SPPRA, Innsbruck, Austria, 2008.

[21] F. H. Zawaideh," Arabic Hand Written Character Recognition Using Modified Multi- Neural Network", Journal of Emerging Trends in Computing and Information Sciences, Vol. 3, NO, ISSN 2079-8407, 7, July 2012.

[22] G. R. Ball and S. N. Srihari, "Semi-supervised Learning for Handwriting Recognition", International Conference on Document Analysis and Recognition, IEEE DOI 10.1109/ICDAR, 249, 2009.

[23] A. Cheung, M. Bennamoun and N.W. Bergmann, "An Arabic optical character recognition system using recognition-based segmentation" ,Pattern Recognition 34 215}233. 2001.

[24] S. A. Mahmoud and A. S. Mahmoud, "Arabic Character Recognition using Modified Fourier Spectrum (MFS) ", IEEE, Proceedings of the Geometric Modeling and Imaging, New Trends (GMAI'06) 0-7695-2604-7/06, 2006.

[25] O. Hachour, "The Combination of Fuzzy Logic and Expert System for Arabic Character Recognition", 3rd International IEEE Conference Intelligent Systems, September 2006.

[26] M. Albakoor, K. Saeed and F. Sukkar, "Intelligent System for Arabic Character Recognition", IEEE, 978-1-4244-5612-3/09, 2009.

[27] Marwan Ali.H. Omer and Shi Long Ma," Online Arabic Handwriting Character Recognition Using Matching Algorithm", IEEE, 978-1-4244-5586-7/10, 2010.

[28] A.M. ALlMl, "A Neuro-Fuzzy Approach to Recognize Arabic Handwritten Characters", IEEE, 0-7803-4122-8/97, 1997.

[29] M. Sarfraz1 and A. T. A. Al-Awami, "Arabic Character Recognition Using Particle Swarm Optimization with Selected and Weighted Moment", IEEE, 1-4244-0779-6/07, 2007 INVARIANTS"

[30] A. Benouareth, A. Ennaji and M. Sellami, "HMMs with Explicit State Duration Applied to Handwritten Arabic Word Recognition", International Conference on Pattern Recognition", IEEE (ICPR'06) 0-7695-2521-0/06, 2006.

[31] N. B. Amor and N. E. Ben Amara," Combining a hybrid Approach for Features Selection and Hidden Markov Models in Multifont Arabic Characters Recognition", IEEE Second International Conference on Document Image Analysis for Libraries (DIAL'06) 0-7695-2531-8/06, 2006.

[32] N. Ben Amor, M. Zarai, and N. E. Ben Amara", Neuro-Fuzzy approach in the recognition of Arabic Characters ", 0-7803-9521-2/06/$20.00 §2006 IEEE.

[33] A. Zidouri, "PCA-based Arabic Character Feature Extraction ", IEEE, 1-4244-0779- 6/07, 2007.

[34] H. A. Al Hamad," Use an Efficient Neural Network to Improve the Arabic Handwriting Recognition", International Conference on Systems, Control, Signal Processing and Informatics, 2013.

[35] A. Elbaati, Houcine Boubaker, Monji Kherallah, Adel M. Alimi," Arabic Handwriting Recognition Using Restored Stroke Chronology", International Conference on Document Analysis and Recognition, 2009 10th.

[36] A. M. ALIMI, "An Evolutionary Neuro-Fuzzy Approach to Recognize On-Line Arabic Handwriting", IEEE, 0-8 186-7898-4/97, 1997.

[37] M. N., Ms. Student and K. Faez,"Recognition of Multi-font Farsi / Arabic Characters Using a Fuzzy Neural Network", LEEE, 0-7803-3679-8/96, 1996.

[38] J. Tohka, Introduction to Pattern Recognition, Institute of Signal Processing Tampere University of Technology, www.tut.fi/ wohtopr.pdf, downloaded 12April 2005.

[39] M.T.SABRI and A.MAHMOUD, "Offline Arabic Handwritten Text Recognition: A Survey", ACM Computing Surveys, Vol. 45, No. 2, Article 23, February 2013.

[40] A. Khawaja, S. Tingzhi, N. M. Memon and A. Rajpa, "Recognition of printed Chinese characters by using Neural Network" , IEEE 1-4244-0794-X/06, pp 169-172. 2006.

[41] K. S. Siddharth, M. Jangid, R. Dhir, R. Rani, "Handwritten Gurmukhi Character Recognition Using Statistical and Background Directional Distribution Features", International Journal on Computer Science and Engineering (0975-3397), Vol. 3 No. 6 June 2011.

[42] P. Jhajj, D. Sharma, "Recognition of Isolated Handwritten Characters in Gurmukhi Script", International Journal of Computer Applications (0975-8887), Vol. 4, No. 8, 2010.

