Holistic Arabic Handwritten Word Segment Recognition Using Multi-Level Neural Network

Osama Nayel Al Sayaydeh, Ahmad Bisher

1 MSc, Computer Science Department, Prince Hussein bin Abdullah of Information Technology, Al al-Bayt University
2 MSc, Department of Computer Engineering, Yarmouk University
1 osamasayaydeh@yahoo.com, 2 ahmad.bisher@hotmail.com

ABSTRACT

The increasing demand on digitization of human activities accompanied with developments in interactive technologies between human and computer, have led to an increase in interest of research for those who are focusing in the field of handwritten recognition of characters, words, sentences, and whole documents. The recognition task involves complex processes in artificial intelligence, image and signal processing. Semitic languages are different from European languages in many aspects including complex linguistic structure, implicit characters and concatenation, writing styles, fonts, and writing direction. The Arabic language, as one of the Semitic languages, has many unique characteristics that make the job of recognition even more challenging. Intensive research has been carried out in the recognition of handwritten English; however less effort has been paid for the recognition of handwritten Arabic. To this end, we propose in this paper applying a multi-level neural network for the holistic recognition of Arabic handwritten documents. In the presented methodology, many morphological operations are being applied followed by special objects recognition to localize and process handwritten words before matching them. The goal is to achieve high recognition ratios in short time. Therefore, the proposed recognition approach will be compared with the state of the art techniques in terms of accuracy and recognition speed.

Keywords: Handwritten, Recognition, Neural Networks, Holistic word

1. INTRODUCTION

In spite of the great development in computers, in terms of speed in completing the business and access of information, our primary ambition is to search for computers, able to communicate with human beings in an actual way and also have the ability to distinguish patterns such as voices and handwritten effectively.

Handwriting has received an attention by many researchers and is considered not only one of the most important challenges, but also considered one of the oldest challenge in the field of computer. The reason of this great interest is to improve the communication between human and computer, which makes the computer friendlier to use [1]. Understanding the handwritings by computer is due to the importance of optical character recognition (OCR) for business, office, mailing address, and reading in addition to check processing etc. [2, 3, 4].

A very serious difficulty facing the researchers in recognizing handwriting is the variety and uncertainty of human writing, not only because of the great variety in the shape, but also because of the overlapping and the cursive of the character. The cursive problem not only exists in English but also in Arabic language. In English, the problem is only in handwriting style, but in Arabic the problem is in handwriting and printed styles. For this reason, Arabic is named fully cursive language. Figure (1) shows the cursive problem in Arabic and English languages. For example, figures (1.a, 1.b, 1.d) show the cursive among the characters, but in figure (1.c) there are no cursive characters.

There are two kinds of handwriting recognition: online recognition and offline recognition. In online recognition, the text is entered in the computer using special devices, such as light pen, tablet PC, or PDA, where a sensor traces the pen movements as well as pen-up / pen-down switching and the text is recognized in real time [5, 6]. Offline recognition deals with the text in the form of an image that is incorporated into a computer using a scanner or camera and can be converted to a text that can be modified. By comparing the two methods, offline and online, it was found out that the online way is easier than offline one. This could be due to the fact that in online way we have the preliminary information about the entered text like the movement of pen and (x, y)
coordinates pairs. But in the offline way, we don't have any information about the entered text beforehand, because we are dealing with the form of image [7, 4]. For this reason, the results obtained by using offline are still low when compared with the results of online. In this work, we have used the offline method for recognizing Arabic handwriting.

The offline method has been used in several languages such as Japanese, Latin, Chinese, as well as Arabic. As a result, good findings were obtained in recognizing these languages [8, 9]. Although the Arabic characters are used in more than one language such as Persian and Urdu, the researches published in Arabic are few when compared with other languages (Chinese, English) and the recognition accuracy is still low. This is due to the nature of Arabic language.

More than 400 million people in Africa and Asia use Arabic language in writing and speaking, but there are also some people using only Arabic characters in writing without pronouncing them such as Persian and Urdu. In addition, Arabic language has many features and characteristics that distinguish it from other languages and this makes the process of recognition very difficult [10].

The features of Arabic language make the recognition process more difficult; example of such features is the existence of different types of fonts.

Recognition of Arabic handwriting is very important, especially in the fields of office automation, processing bank checks and the mailing addresses. The recognition of text means to transform the human writings into machine readable text, but the most difficult problem in Arabic handwritten text recognition is the cursive of handwriting, which makes the segmentation process very challenging. Another major problem is the shape difference of the same character if it is written by different persons. This could be due to the variability of human writings in Arabic. Therefore, it is important to develop a system capable of segmenting Arabic word to characters accurately and improve the character recognition ability.

This paper has been concerned with a robust recognizer based on multi-level feed-forward neural networks. To improve the recognition ability, the suggestion is to construct two different recognizers, one for recognizing isolated Arabic characters, and the other for recognizing the segments. The suggested system is tested using 15 different handwritten documents to validate its results.

2. RELATED WORK

The interest in recognition of Arabic writing started lately compared with other languages, such as Japanese, English and Chinese. The first recognition of Arabic writings came into sight in 1975. However, this recognition process was concerned with Arabic printed text. Afterwards, the concern in Arabic has increased through publishing various researches which are not only concerned with printed texts but also interested in handwritten ones. Moreover, there are also researches dealing with the process of dividing Arabic writing (printed or handwritten) into characters or connected parts.

In this paper, important researches were cited concerning the recognition of Arabic language characters, indicating the used methods, the weaknesses, and the conclusions of these researches.

In (2003), Mohammad Sarfraz, Syed Nazin Nawan and Abdul Aziz Al-Khuraidly [11] proposed a technique for recognizing Arabic printing text using neural network. They used four stages for recognizing text: the first stage is pre-processing which includes Binarization, smoothing and normalization. The second stage is segmentation of text into line, word and then into characters. The third stage is features extraction using moment invariant (moments of an image can be thought as the decomposition of image into a series of numbers that describe the distribution of the image function) as features vector. The last stage is neural network which is used for recognition. The system showed recognition rate of about 73%. It is clear that this study did not deal with handwritten Arabic text.

Labiba Souici and Mokhatar Sellami [12], developed an Arabic literal amount recognition system that uses a neuron-symbolic classifier. This system consists of four main stages: the first stage is concerned with structural features extracted from words contained in the amount vocabulary such as (descenders, ascender, loop, dots). The second stage is to build symbolic knowledge base that reflects a classification of words according to their features. The third stage is to use translation algorithm to convert symbolic representation into neural network to determine the neural network architecture and initialize its connection with specific value. The fourth stage is empirical learning used to recognize new handwritten amounts. This study achieved recognition rate 93%. This empirical study was limited to the words literal Arabic amounts such as "المالون", "سبعون".

In (2006), Somay Alm’addeed [Alm 06] proposed a system for recognizing handwritten Arabic words using neural network. The first stage is preprocessing which contains normalization, which attempts to remove some of the variation in the images which do not affect the identity of the word. The second stage is feature extraction. This paper deals with global features (descenders, ascenders, number of loops, number of segments, lower dots, upper dots,) which are extracted from word. The neural network used global feature as input to recognize Arabic word. The final system produced had an overall recognition rate 63%.

A holistic technique for classifying and retrieving Arabic handwritten text document proposed by Salma Brook, Zaher Al-Aghbari in (2008) [14]. The system consists of three main steps: first is segmenting the Arabic text images to words; then, the words images are segmented to connected parts. Second, several features are extracted from these connected parts and then
combined to represent a word with one consolidated feature vector. Finally, classification step using feedforward Neural Network is used to classify the connected parts.

In the same year, Z. shaaban [15], suggested a new approach to tackle the problem of recognizing machine printed Arabic texts. The first step in the system is normalization of scanned image for reducing the size of image, and then segmented the normalized image into characters, finally, sent the segmented image to classifier. The suggested scheme depends on multiple parallel neural networks classifier. It consists of two phases. First phase categorizes the input character into one of eight groups. Second phase classifies the character into one of the Arabic character classes in the group. The system has been tested on more than 100 Arabic text images. The system achieved high recognition rate 98%. This study dealt with the printed Arabic text only.

In (2009), Jawad H. Alkhatteeb, Jianmin Jiang, Jinchang Ren and Stan S Ipson [16] proposed a machine learning approach for classifying handwritten Arabic word. The proposed approach consists of three stages. Pre-processing stage, consists of word segmentation and normalization to remove as much as possible of the variations in handwritten images for consistent analysis and robust recognition. Feature extraction stage used three different feature extraction methods for each segmented word namely the Discrete Cosine Transform (DCT), Moment Invariants, and Absolute Mean Value of overlapping blocks. Finally, classification stage used the features that are extracted in second stage to train NN to recognize words. The system used IFN/ENIT database for testing purposes which consist of 32492 Arabic words. The recognition rate from using DCT features is 80.74% and from using Moment Invariants are 75.74.

3. METHODOLOGY

The presented algorithm consists of two phases; preprocessing and recognition. Figure (2) shows the block diagram of the presented algorithm.

![Figure 2: Presented system block diagram](image1)

The assumption of this paper is that, the input is a segmented Arabic handwritten word segment. This segment could be consisting of one character or more than one character. A special neural network is being design to recognize the single character segments, and another separate neural network was designed to recognize multi-character word segment.

Preprocessing has been conducted so as to get rid of noise available in the image, in addition to do morphological operations for feature extraction before recognition. The result of feature extraction those are being gotten by the preprocessing are the input for the neural network recognizers.

Two levels of recognizer are implemented in the design. Level one consists of two neural networks, and level two is consisting of one neural network that is function is decision making. In decision level, a comparison of the output of characters of neural network with the output of segments of neural network has been made. Figure (3) illustrates the multi-layer neural networks propagation.

![Figure 3: Multi-level neural network connection](image2)
it transforms the input image F to a binary image G depending on threshold T.

The goal of smoothing process is to reduce the noise (noise may be bumps in edge, or small gaps) from the scanned document images. Smoothing can fill the small gaps, or remove the small bumps in edges. In this paper, the morphological operations have been used. The morphology consists of two main operations: dilation and erosion [17]. Generally, dilation operation adds pixels to the bounders of object in an image, while the erosion operates reversibly. The number of pixel added or removed depends on the size of structure elements that is used in the design. The paper uses the first operation dilation, which is represented in the equation (1) [18, 19]:

\[ A \ominus B = \{z \in (B_x) \cap A \neq \emptyset\} \]  (1)

B is a set of structure elements where \( \emptyset \) is the empty set and B is structure elements. In other words, the dilation of A by the use of B is the set consisting of all structure elements origin locations where the reflected and translated B overlaps at least one some of portion of A. Two types of dilation are applied: horizontal structure zones, and vertical structure zones. The result of applying dilation on a binary image is illustrated in figure (4).

**Figure 4:** Illustration of binary word dilation; original image (a), dilated image (b)

Feature extraction is the process of extracting useful information from image. A number of properties of the binary dilated image will be extracted. These properties will be used for as input to the recognition neural network in addition to the holistic word segment.

The extracted properties are centroid, area, and the ratio of white pixels with respect to black pixels.

The area is the number of pixels within the boundary box that bounds the word segment. The area of each region within image is calculated by counting the number of pixels in the. This is done by the equation (2) [20]:

\[ Area = \sum_i \sum_j f[i,j] \]  (2)

Where \( f[i,j] \) represents the pixels of labeled region.

The centroid or (center of mass) is the point that is used to specify its position of segment center. It could be calculated by the two equations (3, and 4) [20] with respect to x-axis and y-axis respectively:

\[ \bar{X} = \frac{1}{A} \sum x_i \]  (3)

\[ \bar{Y} = \frac{1}{A} \sum y_i \]  (4)

Where \( x_i \) and \( y_i \) is the coordinate of points in the region;

\( \bar{X} \) and \( \bar{Y} \) are the region Centroid; A is the area of the region.

The recognition phase consists of two levels; neural networks to recognize the fragments and characters respectively, and decision making level.

Neural networks will be used for carrying out the task of recognition because of their efficiency and high ability to recognize the patterns. In this phase, the single characters, and connected-parts segments that have been segmented previously, will be recognized using two suggested neural networks. The first neural network is trained for recognizing the Arabic characters only, whereas the second neural network is trained to recognize the fragments that are difficult to be segmented to characters due to the quality of ligature.

Once an image has been segmented, it will be fed to both neural networks at the same time in order to be recognized. Then, a decision will be made (if it should be recognized as character or as fragment) according to a declared rule. The output of the second level neural network determines that, the input is either character or segment or none of them.

Multi-layer perceptron (MLP) neural network were designed, trained, and used in this paper. The two suggested neural nets have different architectures (input, hidden, and output nodes), each is trained with its own parameters, and different training set.

The image resulted from the segmentation phase is 40×70 pixels. The two nets have different numbers of input nodes (the character recognizer net has 36 input node, while the fragments recognizer net has 2800 node in the input layer). Since the resulted image is fed to both neural nets, the segmented image will be resized to 6×6 pixels before feeding it to the character recognizer neural net.

The learning rate of the first level recognition neural network is 0.2 for both networks, while it for the second level neural network is 0.01. This was due to the high similarity in input vectors for different classes. The learning rate determines how the weights and biases could be changes in maximum change every trial of learning. The performance goal was selected for all training process to be sum square error.
The first level neural networks structures are illustrated in figures 5 and figure 6. Each vertical nodes set represents a neural network layer which has its own transfer function, weights vector and bias vector.

![Figure 5](image)

**Figure 5:** Structure of neural network for single character recognition

![Figure 6](image)

**Figure 6:** Structure of neural network for multi-character segments recognition

The decision is made to be based on the following rules:

- Step1: If the output of the character recognition neural net is unknown, check the fragments recognizer. The segment is recognized based on neuron with the maximum value. (i.e. it is recognized as the section corresponding the maximum neuron).

- Step2: If the output of the character recognition neural net is a number n between (1-28), also check the output of the fragments recognizer.

- If the maximum value output neuron is < Tn2 (Tn2 = 0.2, form the fragments neural) then, the decision is the character corresponding to the character that n represents. Else, the segment is recognized as unknown.

4. RESULTS

The results obtained those are gotten by implementation of the presented system before are going to be presented here. To study the behavior of the presented system, two types of data have been collected. One has been only assigned to training and testing the neural network associated with characters, and to investigate the aspects of strength and weakness of recognition process. And the third one has been used to train the neural network concerning recognizing fragments to investigate the strengths and weaknesses of the proposed system. Here the data has been divided into training data and testing data.

The data was collected from 15 people. Each person wrote 15 Arabic documents by his / her hand. Each handwritten document is different from other handwritten documents in size and in the way of handwriting. Moreover, the words of handwritten documents are different. Each document consists of 5 lines. Each line has about 10 or 12 words, thus, getting about 50-60 words in each document. The total number of the handwritten words collected from the subjects is nearly 800 words. Figure (7) shows an example of the collected dataset.

![Figure 7](image)

**Figure 7:** Sample of collected data

The data set is divided into training set and testing set. The training set that represents 90% of total data set and it was assigned to training the neural network on different shapes of characters. The reason for this is that Arabic characters have different shapes in handwriting; each person wrote characters differently and even the same person wrote the characters differently.

The testing set represents 10% of the total data set. It was assigned to testing the ability of neural network to recognize the characters. Each image in the data set represents a character in different shapes.

The ability of neural network to recognize each character of handwritten Arabic characters is shown as the result of recognition in table (1). Hence, the Arabic language is structured from 28 characters; the table illustrates each character recognition accuracy percentage separately.
It is noticed that the highest recognition ratio was for (أ). This can be attributed to the non-existence of a character similar to (أ) shape. For the other remaining Arabic characters, the ratio ranged between 68% - 80%. This is due to the existence of similarity among these characters in spite of the difference in the position of dot.

### Table 1: Recognition result of Arabic single characters

<table>
<thead>
<tr>
<th>Characters</th>
<th>Recognition ratio %</th>
</tr>
</thead>
<tbody>
<tr>
<td>أ</td>
<td>88</td>
</tr>
<tr>
<td>ب</td>
<td>74</td>
</tr>
<tr>
<td>ت</td>
<td>70</td>
</tr>
<tr>
<td>ث</td>
<td>82</td>
</tr>
<tr>
<td>ج</td>
<td>68</td>
</tr>
<tr>
<td>ح</td>
<td>78</td>
</tr>
<tr>
<td>خ</td>
<td>70</td>
</tr>
<tr>
<td>د</td>
<td>76</td>
</tr>
<tr>
<td>ر</td>
<td>64</td>
</tr>
<tr>
<td>ز</td>
<td>84</td>
</tr>
<tr>
<td>س</td>
<td>76</td>
</tr>
<tr>
<td>ش</td>
<td>78</td>
</tr>
<tr>
<td>ض</td>
<td>74</td>
</tr>
<tr>
<td>ط</td>
<td>73</td>
</tr>
<tr>
<td>ظ</td>
<td>70</td>
</tr>
<tr>
<td>ع</td>
<td>50</td>
</tr>
<tr>
<td>غ</td>
<td>66</td>
</tr>
<tr>
<td>ف</td>
<td>77</td>
</tr>
<tr>
<td>ق</td>
<td>73</td>
</tr>
<tr>
<td>ك</td>
<td>62</td>
</tr>
<tr>
<td>ل</td>
<td>80</td>
</tr>
<tr>
<td>م</td>
<td>77</td>
</tr>
<tr>
<td>ن</td>
<td>74</td>
</tr>
<tr>
<td>ه</td>
<td>78</td>
</tr>
<tr>
<td>و</td>
<td>68</td>
</tr>
<tr>
<td>ي</td>
<td>74</td>
</tr>
</tbody>
</table>

**Average Ratio**: 72.8%.

The average ratio of recognizing Arabic characters reaches 72.8% as illustrated from table (1).

Table (2) shows the testing result of the segments recognition. It is clearly seen that the highest ratio of recognition was for the fragment (فِي) because of the non-existence of fragment similar to this fragment and the uniqueness of this structure shape. On the other hand, the lowest ratio was between the two fragments (مـٓحـٓ). This is because of the existence of similarity between these two fragments.

### Table 2: Result of the Recognition of fragments

<table>
<thead>
<tr>
<th>Connected parts</th>
<th>Recognition rate%</th>
</tr>
</thead>
<tbody>
<tr>
<td>الفين</td>
<td>80</td>
</tr>
<tr>
<td>المـٓحـٓ</td>
<td>72</td>
</tr>
<tr>
<td>المـٓحـٓ</td>
<td>70</td>
</tr>
<tr>
<td>المـٓحـٓ</td>
<td>85</td>
</tr>
</tbody>
</table>

**Ratio**: 76.75%.

After having declared the recognition ratio of each fragment, the total ratio of recognizing fragment is 76.75%.

### 5. CONCLUSION

The total ratio of both neural nets for recognizing both isolated characters and segments reached about 74.77%. The resulted recognition ratio is very good.

Generally, when comparing the ratio of recognition that the applied system has reached with those of other systems, it can be said that it achieved the best ratio in some cases. For instance, when compared with Muhammad Sarfraz and et-al [11] who had total recognition ratio 73%, it can be concluded that the presented system has got better results. Moreover, when compared our system with Somay Alm'addeed [7] who developed a system for recognizing handwritten words and got about ratio 63%, it has become obvious that our results exceed the results of her study by about 11%. This is due to the weakness of her system which was based on the global features for recognizing Arabic handwritten words. In fact, this is very difficult since it couldn’t rely on these global features for recognizing the handwritten words because handwriting varies from one person to another. On the contrary, if these features were applied to one type of printed fonts, the results would be better.

### REFERENCES


IEEE/ACS International Conference on computer systems and application, 2009, pp. 582-590


