International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol. 8, No. 2 (2015), pp. 401-426
http://dx.doi.org/10.14257/ijsip.2015.8.2.37

# A Survey on Arabic Character Recognition

Ahmed Lawgali

*University of Benghazi*
*Benghazi - Libya*
*Ahmed.lawgali@uob.edu.ly*

***Abstract***

*Off-line recognition of text play a significant role in several application such as the automatic sorting of postal mail or editing old documents. It is the ability of the computer to distinguish characters and words. Automatic off-line recognition of text can be divided into the recognition of printed and handwritten characters. Off-line Arabic handwriting recognition still faces great challenges. This paper provides a survey of Arabic character recognition systems which are classified into the character recognition categories: printed and handwritten. Also, it examines the literature on the most significant work in handwritten text recognition without segmentation and discusses algorithms which split the words into characters.*

***Keywords:*** *Arabic character recognition and character segmentation*

## 1. Introduction

Automatic recognition of content of a document is the ability of the computer to distinguish the differences between individual characters and words. Optical character recognition is the transferring of the image of the text into an editable text to avoid retyping it. Such recognition could be used in various applications, for example: the automatic sorting of postal mail, cheque processing or editing old documents. Character recognition can be divided into two types: online and off-line [1]. In online recognition, characters are recognised during the writing process using the digitised trace of the pen. However, off-line recognition deals with images scanned of previously written documents. Many languages such as Persian, Urdu and Jawi, use Arabic characters [2]. Nevertheless, compared to Latin and Chinese handwritten character recognition, little research has been done into Arabic handwritten character recognition.

The organization of the paper is as follows. Section [2] discusses the challenges and motivation for Arabic character recognition. Section [3] provides a background to off-line handwritten text recognition while Section [4] provides a literature review on Arabic character recognition which is classified into the character recognition categories: printed and handwritten. Section [5] discusses directions for future work. Section [6] concludes the paper.

## 2. Challenges and Motivation

Arabic script is written from right to left and is composed of 28 characters, with no upper or lower case. Each character has two or four shapes; the shape of the character depends on its position in the word, as shown in Table 1. The first column gives the number of the character, the second column is its name, the third represents the sign of an isolated character, with the fourth being its appearance at the beginning of the word.

ISSN: 2005-4254 IJSIP

**Table 1. Arabic Character Forms**

| No | Name | Isolate | Beginning | Middle | End |
|---|---|---|---|---|---|
| 1 | Alif | ا | - | - | ـا |
| 2 | Baa | ب | بـ | ـبـ | ـب |
| 3 | Taa | ت | تـ | ـتـ | ـت |
| 4 | Thaa | ث | ثـ | ـثـ | ـث |
| 5 | Jeem | ج | جـ | ـجـ | ـج |
| 6 | Haa | ح | حـ | ـحـ | ـح |
| 7 | Khaa | خ | خـ | ـخـ | ـخ |
| 8 | Daal | د | - | - | ـد |
| 9 | Dhal | ذ | - | - | ـذ |
| 10 | Raa | ر | - | - | ـر |
| 11 | Zaa | ز | - | - | ـز |
| 12 | Seen | س | سـ | ـسـ | ـس |
| 13 | Sheen | ش | شـ | ـشـ | ـش |
| 14 | Saad | ص | صـ | ـصـ | ـص |
| 15 | Dhad | ض | ضـ | ـضـ | ـض |
| 16 | Tta | ط | ط | ـطـ | ـط |
| 17 | Dha | ظ | ظ | ـظـ | ـظ |
| 18 | Ain | ع | عـ | ـعـ | ـع |
| 19 | Ghain | غ | غـ | ـغـ | ـغ |
| 20 | Faa | ف | فـ | ـفـ | ـف |
| 21 | Qaf | ق | قـ | ـقـ | ـق |
| 22 | Kaaf | ك | كـ | ـكـ | ـك |
| 23 | Laam | ل | لـ | ـلـ | ـل |
| 24 | Meem | م | مـ | ـمـ | ـم |
| 25 | Noon | ن | نـ | ـنـ | ـن |
| 26 | Haa | ه | هـ | ـهـ | ـه |
| 27 | Waaw | و | - | - | ـو |
| 28 | Yaa | ي | يـ | ـيـ | ـي |

Finally, the fifth and sixth columns represent its appearance in the middle and at the end of the word, respectively. The dots play a significant role in Arabic characters. The shape of some characters is similar but the difference arises with position and number of dots, this can occur either above or below the characters. For example, three characters such as (ب, ت, ث) have a similar shape. Dots may appear as two distinct dots or may be connected into a line in handwritten texts. Furthermore, short marks such as a "hamza", can be placed above or below five particular characters or can appear as isolated characters. Some Arabic characters have a loop, such as (ص, ف, و). Moreover, Arabic text is cursive; meaning that characters of a word are connected through an imaginary horizontal line called a *baseline*. Also, there are lines which appear above and below baseline, called *ascender* and *descenders*, as shown in Figure 1 [3].

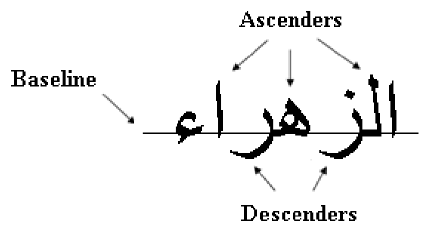
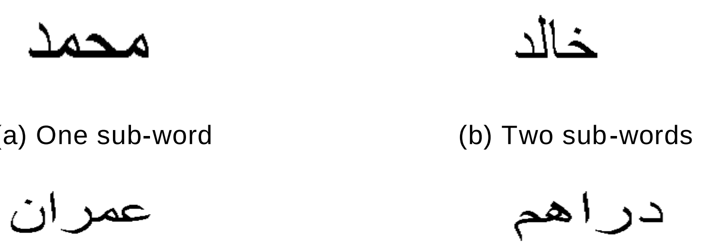



**Figure 1. Baseline, Ascenders and Descenders as Shown in a Word**

In addition, there are six characters (ا, د, ذ, ر, ز, و) that have not shapes at the beginning and in the middle of the word, as shown in Table 1. Therefore, these characters do not connect to a subsequent character in a word and this causes a separation of the word into parts, as shown in Figure 2. These parts are called *sub-words*.




(a) One sub-word (b) Two sub-words

(c) Three sub-words (d) Four sub-words

**Figure 2. Words with Different Sub-words**

Spaces separate words and short spaces separate sub-words. The same word in a handwritten Arabic text could have different styles and sizes for the same writer as well as for different writers, as shown in Figure 3.

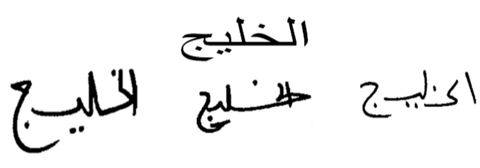



**Figure 3. The Same Word Could Have Different Styles and Sizes**

Also, two or more characters in the handwritten Arabic language can be combined vertically and represented by different shapes. This overlap between the neighbouring characters is called a *ligature* and means that the second character might appear before the first one in some cases [83]. A ligature might occur when characters such as (ح, خ, ج, م, ل) appear after certain other characters. Figure 4 illustrates the overlapping in the first three characters(خ, ل, ا). Moreover, two characters can be vertically overlapped without touching one another, such as characters(ر, ا, ع), in Figure 5. In some cases, two characters may touch by mistake, such as characters (و, ر), in Figure 6.

**Figure 4. Overlapping in Arabic Characters**

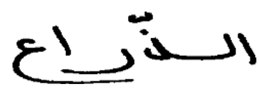


**Figure 5. Overlapping without Touching in Arabic Characters**

**Figure 6. Touching by Mistake in Arabic Characters**

In handwritten Arabic texts, some characters appear to be similar, although they are different, and it is difficult even for the human eye to find the difference [18]. There are differences between the length and the width of Arabic characters, for example (ا, ب). Also, the same character may appear different in its various forms, such as (ع, عـ) [113]. Moreover, the great similarity between some of the handwritten characters makes classification these characters another challenge, as shown in Figure 7.

**Figure 7. Similarity between Some of the Characters**

## 3. Arabic Character Recognition System

Off-line handwritten character recognition involves numerous challenges owing to the complexity and ambiguity in styles of writing. Character recognition systems can be classified into different categories, as shown in Figure 8.

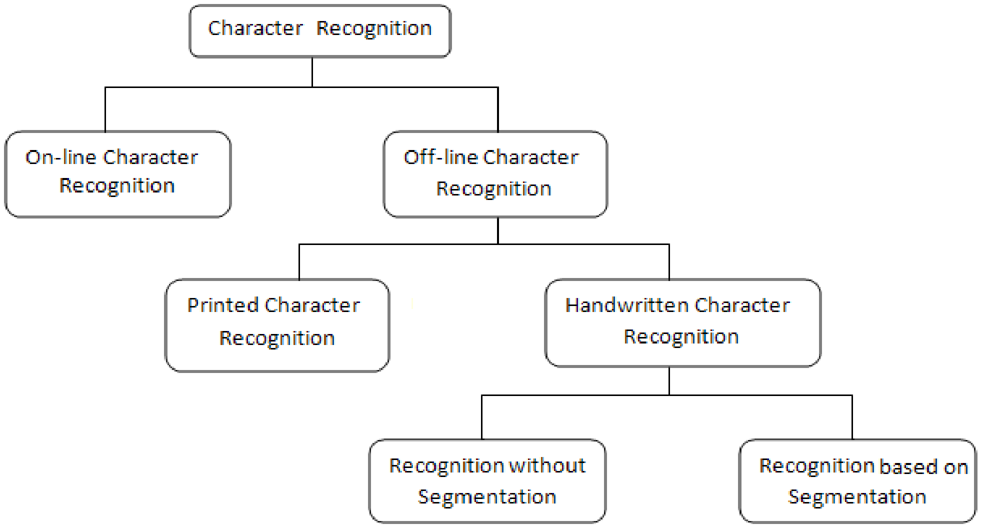



**Figure 8. Classification of Character Recognition Systems**

The Figure indicates different routes for Arabic character recognition with or without segmentation. Character recognition can be divided into two types: online and off-line character recognition [6]. The online character recognition is classified while the user is writing (see [7, 8]). This technique employs equipment such as a special pen and a tablet, the digitised trace of the pen being used to recognize the character. Therefore it could not be applied for recognising pre-written documents. For further details about the difference between the online and the off-line character recognition, readers are referred to the paper [9]. On the other hand, the off-line recognition systems deal with images scanned of previously written documents. The off-line recognition of texts can be further divided into two categories: the recognition of printed and of handwritten characters. Printed characters have one style and size for any given font. However, handwritten characters have styles and sizes which vary, both for the same writer and between different writers. Handwritten words can be recognised in two ways: recognition of a whole word without segmentation, or recognition based on segmentation. Due to the presence of the ligature and to the cursive nature of Arabic script, several researchers have presented techniques based on recognition of the whole word without segmentation [10, 11]. Handwritten word recognition involves several steps to achieve classification as a text file. Figure 9 illustrates the general steps in handwritten text recognition. Because the segmentation process is the main source of errors in recognition, most systems avoid this step and merely recognize the whole word.

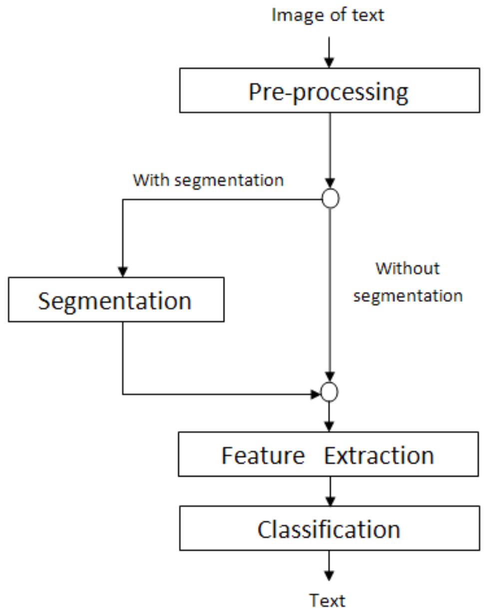



**Figure 9. General Steps for Handwritten Text Recognition**

### 3.1. Pre-processing

The pre-processing task is an important step in any recognition system [12]. The purpose of this step in handwritten text recognition is to enhance the readability of the text image and remove those details that have no discriminative power in the process of recognition [13, 14]. Mowlaei, *et al.*, in [15] normalised characters into size *64x*64 pixel image and skeletonised the stroke width. Cheung, *et al.*, in [16] presented a simple smoothing method to reduce the noise. Nawaz, et *al.*, in [17] converted the image to monochrome in order to remove noise and Drift correction is used for skew detection and correction. Altuwaijri and Bayoumi in [18] achieved binarisation and smoothing the image by eliminating the noise. The pre-processing step usually includes several tasks: binarisation, noise removal, baseline detection and normalisation.

**3.1.1. Binarisation and Noise Removal:** Binarisation is a process that is used to convert a text image into a binary format: values of background pixels as 1 (white) and values of foreground pixels as 0 (black). This process is carried out by choosing an efficient thresholding method value. One of the advantages of this process is that it increases the processing speed [19]. Distortion in the image usually occurs during the scanning process. Small objects not part of the writing can be considered as noise and need to be removed. There are some methods used to reduce the noise, these methods include filtering and morphological operations [20]. Many spatial and frequency domain filters can be designed for different purposes, such as smoothing [21] and removing noise [22]. Smoothing is reducing the noise by using mathematical morphology operations. It is usually achieved in two ways: by opening the spaces between touching objects or by filling small gaps [23]. Both methods, opening and closing, apply the morphology operations of erosion and dilation respectively. Mathematically [24], the erosion of *A* by *B* is the set of all points *z* and can be defined as

$$A \ominus B = \{z \mid (B)_z \subseteq A\}$$

and the dilation of *A* by *B* defined as

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \emptyset\}$$

The morphological operations can be designed for various purposes such as smoothing the contours [25], connecting the broken strokes [26] and extracting the boundaries [27].

**3.1.2. Baseline Detection:** The baseline is an imaginary horizontal line that connects the characters of a word [3]. Detecting the baseline is one of the important tasks in the pre-processing stage [14]. It could be a useful in skew/slant correction as well as character segmentation [28]. A common method used for extracting the baseline uses a horizontal projection and this works well, especially with printed text [28, 29]. Detecting the baseline helps to determine certain structural features, such as dots and their positions, ascenders and descenders, as well as enabling skew/slant correction. There are different methods applied to detect the baseline. The horizontal projection method was used for detecting the baseline by various authors (Parhami and Taraghi, [30]; Nawaz, *et al.*, [17]; El-Hajj, *et al.*, [11]; Al-Rashaideh, [13]). The method based on the word skeleton to detect Arabic handwriting baseline was proposed by Pechwitz and Margner [29]. However, Farooq, *et al.,* in [14] proposed another method based on the word contour. Burrow in [31] proposed a method based on Principal Components Analysis (PCA) to detect Arabic baseline. This method detected baseline direction based on the distribution of foreground or background pixels.

**3.1.3. Normalisation:** Handwriting has different styles and sizes. Therefore, the normalisation process is one of the most important tasks in the text recognition process. It is used to reduce the variation between the images of the text and to adjust the size of the character or the word. Size normalisation is commonly used to reduce size variation and adjust the character or word sizes to an identical size [32]. Some researchers such as Yanikoglu and Sandon in [33] proposed dividing the character into a number of zones and then each zone scaling separately. Owing to inaccuracies in the scanning process, the scanned document may be slightly rotated. Skew correction is used to correct the orientation angle. The baseline for the text is usually used to detect the skew in Arabic texts. In 2005, Dong and others proposed an algorithm to correct skew and slant the word, based on Radon transform [34].

### 3.2. Segmentation

The segmentation stage entails segmenting the text into its sub-units, such as *lines*, *words* or *characters*. Segmentation is an important stage because it has an effect on the recognition rate [4, 35]. There are several types of segmentation:

- **Segmenting a Page into Lines:** In general, the paragraph or page consists of several lines. In text recognition, the text has many lines segmenting into separated lines. There are certain methods used for this purpose, such as horizontal projection [9, 36].
- **Segmenting a Line into Words:** After segmenting lines, the line is segmented into words. Segmenting the line into words depends on the space between the words. However, there are some characters which do not connect to a subsequent character in a word and this causes a separation of the word into sub-words. Longer spaces separate words while short spaces separate sub-words. Therefore, most researchers assume in their technique that the space between words is bigger than the space between sub-words [37]. Some of the methods analyse the distances between connected components to segment the words [38].

- **Segmenting a Word into Characters:** This part of segmentation involves segmenting the word into individual characters. As described in [39], the segmentation points are identified at the end of a character and at the beginning of the next one. The cursive nature of the Arabic language makes segmentation of words into individual characters a difficult task.

There are different techniques applied for character segmentation [2, 4]:

**1. Segmentation Technique based on Vertical Projection:** The vertical projection technique reduces two-dimensional information into one dimension. This technique is based on the fact that the thickness of the connection stroke between the characters in the word is always less than the other parts. Many algorithms have been proposed in this technique for character segmentation such as [17, 30, 39, 40]. The vertical projection is useful for segmenting the words, sub-words and characters, while the horizontal projection is usually used for line segmentation and baseline extraction**.**

2. **Segmentation Technique based on Thinning:** In character recognition, the skeleton of a character provides the essential information about the shape of the character. A number of algorithms have been proposed for extracting the skeletons [41, 42, 43]. Tellache, *et al.*, in [42] used the thinned word to identify edge points, end-points and break-points. Al-Sadoun and Amin, in [44] traced the thinned word from right to left to detect segmentation points.

**3. Segmentation Technique based on Contour Tracing:** Another technique based on contour tracing of a word is also used for segmentation. This technique is achieved by tracing the outer contour of the main body of the word. Sari, *et al.*, in [45] proposed an algorithm for character segmentation based on contour analysis and topological rules. It starts by finding the local minima point of the lower contour. Then, it uses the topological rules to determine whether the local minima point is a segmentation point.

4. **Segmentation Technique based on Artificial Neural Networks:** ANNs are used to verify the valid segmentation points. In [46], Hamid and Haraty proposed a technique for segmenting handwritten Arabic texts based on ANNs. They identified pre-segmentation points by using topographic features of each connected block of characters, such as black pixel density and holes. The ANNs are then used to verify these segmentation points. The potential points are manually classified into valid and invalid segmentation points, and these points are fed with their features, into ANNs for training [2].

5. **Segmentation Technique based on Morphological Operators:** Morphological operations such as closing followed by opening are used for character segmentation. Motawa, *et al.*, in [47] have proposed an algorithm for segmenting handwritten Arabic words. They assume that the vertical strokes which might represent the start and end of the character are found by singularities. Regularities that have information are required to connect a character to a subsequent character. Therefore, the original image is scanned from right to left in order to identify the regularities. Singularities are identified by applying an opening to the word. The regularities are extracted by subtracting the singularities from the word image. These regularities are used to identify the candidates of segmentation points.

However, most of the segmentation algorithms currently proposed does not solve the problem of overlapping characters in Arabic handwriting. The segmentation stage is the most difficult stage and the main source of errors in recognition. Segmentation techniques therefore still represent a challenge in text recognition and need to be improved [2].

### 3.3. Feature Extraction

In printed and handwritten texts, the features capture the information extracted from the text image. This information should have the essential characteristics of the character or the word which make it different from another; in other words, filtering out all attributes and preserving the properties that make one character or word different from another [23]. This information is passed onto the classifier to assist in the classification process. Feature extraction techniques differ from one application to another. Techniques that succeed in one application, may not be successful for other applications [48]. Therefore, the selection of the method for feature extraction remains the most important step for achieving high recognition accuracy. Features of handwritten texts can be classified into the following categories:

**3.3.1. Structural Features:** Structural features are the most common features used by researchers [49]. They illustrate the geometrical and topological characteristics of the text image by describing their local and global properties [23]. Structural features depend on the category of the pattern to be classified. For an Arabic text, for example, the features include dots and their position, strokes, width and height of the stroke, directions, intersection of line segments and loops [50, 51].

**3.3.2. Statistical Features**: Statistical features analyse the spatial distribution of pixels by counting local features at each pixel and deriving a set of statistics from the distributions of the local features [23, 52]. The major statistical features of the character include zoning, where the character divides into overlapping or nonoverlapping zones and the density distribution of character pixels in different regions is analysed. Mohiuddin and Mao in [53] measured the direction of the contour of the character by dividing the image character into zones. Then, histograms of chain codes are used to compute contour direction in these regions.

**3.3.3. Global Transformation**: Global transformation techniques have the ability to convert pixel representation into a more compact form. These techniques can represent the signal by a linear combination of a series of simpler well defined functions. The series expansion provides a compact encoding by the coefficients of the linear combination [19, 20]. Common transformation techniques used in character recognition are:

- Fourier Transform [54].
- Discrete Cosine Transform [10].
- Wavelets [55].
- Hough Transform [56].
- Moments [17].

### 3.4. Classification

Classification is a task which attempts to identify an object by comparing its features to one of a given set of classes. It assumes that the training classes and the training model have been provided. The training model consists of a set of instances where the right answers (outputs) for every input are provided. Therefore, a classifier is used to identify an object by using its features, these are then compared and saved as models for the classes trained. The features of an object in the testing phase will be extracted and compared with the features of the training models to identify the unknown object. Common methods in the classification stage are based on:

- Artificial Neural Network [57, 58, 59].

- Hidden Markov Model [60, 61, 62].
- K-nearest neighbour [63].

# 4. Literature Review on Arabic Character Recognition

This section reviews previous works on Arabic character recognition systems. Insignificant amount of research has been carried out in the field of Arabic handwritten character recognition compared to research into its Latin and Chinese counterparts [55]. The literature review shows that, by surveying relevant works, Arabic character recognition systems can be classified into different categories, printed Arabic text recognition and handwritten text recognition. To develop a recognition system, a database of characters is needed to facilitate the recognition process. Owing to the lack of a complete and publically available reference database of Arabic characters, most researchers have developed their own systems based on a set of data they themselves have gathered [64]. These databases are not available for public use.

## 4.1. Databases

For a fair comparison of performance of Arabic handwriting systems, a standard database is needed. Some of the handwriting recognition systems yield high accuracy due to results being tested on a small database and not a standard database [19]. Most of the databases have also been developed for specific domains, such as postal addresses and financial use (for example, cheques). In handwritten text, there is no control over writing style or the writer's ability to produce text with touching or overlapping characters. Therefore, to develop a recognition system, ideally, a large database for training and testing is required. As previously stated, most of the research has been done on Latin handwritten character recognition. Therefore, a number of databases for handwritten English recognition have been developed. Hull in [65] developed a handwritten English text database for the Centre of Excellence for Document Analysis and Recognition (CEDAR) containing approximately 5,000 city names, 5,000 state names, l0,000 ZIP Codes, and 50,000 alphanumeric characters. The data were scanned with a quality of 300 dpi and grey scale. The database is divided into two groups for the purposes of training and testing. Another large database which includes English characters (upper and lower cases) as well as number digits was provided by The National Institute of Standards and Technology (NIST) [66]. The MNIST database was developed out of the original NIST [67]. It is divided into two groups (60,000 examples for training and 10,000 examples for testing). The IAM database of handwritten English text is another example provided by Marti and Bunke [68]. This database consists of 82,227 word instances of vocabulary written by 400 different writers. However, there are only a limited number of databases for Arabic handwriting, which have been developed for specific domains. One of the most used is the IFN/ENIT database released by Pechwitz and Maergner [69] for handwritten Arabic words. It contains 26,459 handwritten Arabic names. The IFN/ENIT database was written by 411 different writers to represent 937 Tunisian town/village names. The database is divided into five groups for the purposes of training and testing and it is available free for the purpose of research *(www.ifnenit.com).* The Isolated Farsi Handwritten Character Database (IFHCDB) contains 52,380 isolated characters and 17,740 numerals, presented by Mozaffari, *et al.*, [64]. The data were scanned with a resolution of 300 dpi and the images are stored as *77x95* BMP images. The IFHCDB database is divided into two sets (70\% for training and 30\% for testing). This database is available at *(http://ele.aut.ac.ir/imageproc/downloads/IFHCDB.html)* for the purpose of research. The Arabic handwritten database (AHDB) developed by Alma'adeed, *et al.*, [70, 71] has been written by 100 different writers and contains numbers and quantities used in cheques, as well as the most popular words in Arabic writing. Al-Ohali, *et al.*, in [72] have built an Arabic cheque database containing 29,498 sub-words, 15,175 Indian-Arabic digits and 2,499 samples each of legal and courtesy amounts extracted from 3,000 real-life cheques. The ADBase database has been developed by El-Sherif and Abdleazeem [73] for handwritten Arabic digits and contains 70,000 digits written by 700 different writers where each digit is written ten times. This database is divided into two groups (60,000 words for training and 10,000 words for testing). A modified version of the ADBase called MADBase, has been developed by the same authors [73]. This database has the same format as the Latin database (MNIST) to enable comparison between Latin and Arabic digit recognition problems. Each digit image of the ADBase is normalised to a size of *28x28* pixels. The images of the MADBase are grey-scaled, where the ADBase images were binary. Both the ADBase and the MADBase are available free for the purpose of research from *(http://datacenter.aucegypt.edu/shazeem).* Awaidah and Mahmoud in [74] have developed a database for handwritten Arabic digits containing 21120 samples written by 44 different writers. Each writer wrote 48 samples by using a semitransparent paper over a tabular grid. The samples were scanned with a quality of 300 dpi and were transformed into binary images. For each scanned paper, the horizontal and vertical histogram has been computed to extract the location of each digit. Then, each digit was saved in a file. A database for handwritten Arabic characters containing 15800 character samples written by more than 500 writers has been developed by Asiri and Khorsheed [59]. The data collected on A4 sized forms; each form consists of *7x7* small squares. The writer has to fill one character inside each square. The forms were scanned with a resolution of 300 dpi and each character sample is stored in a file as grey scale image. This database is not available to other researchers. Lawgali, *et al.,* in [75] have developed a database for handwritten Arabic characters (HACDB). This database has designed to cover all shapes of Arabic characters including overlapping ones. It contains 6,600 shapes of characters. The data was collected from 50 writers, their ages ranging from 14 to 50. Each writer filled in two forms all the shapes of Arabic characters as (isolated, at the beginning of the word, in the middle and at the end of the word) also the shapes of the overlapping characters but without dots. The database is divided into two versions. The first version is used for training and testing. This version can be used for purposes such as to compare the effectiveness of some techniques for recognition of all the shapes of Arabic handwritten characters. The second version stores all shapes and can be used for training in approaches which recognize the word after segmentation. The HACDB database is available for academic use at *(http://computing.unn.ac.uk/characterdatabase/).* Table 2 lists the databases available for Arabic character recognition.

**Table 2. The Databases Available for Arabic Character Recognition**

| Author(s) | Database name | Type of data |
|---|---|---|
| Pechwitz *et al.* [69] | IFN/ENIT | Words |
| Mozaffari *et al.* [64] | IFHCDB | Isolated characters and numerals |
| El-Sherif *et al.* [73] | ADBase<br>MADbase | Digits |
| Lawgali *et al.* [75] | HACDB | Shapes of Arabic characters and overlapping ones |

### 4.2. Printed Arabic Text Recognition

Printed characters do not possess the great complexity and ambiguity of handwritten characters because they have one style and size for any given font. In this section, the most significant works on recognition of printed Arabic texts are considered. In [76], Amin and Al-Sadoun put forward a structural technique for automatic recognition of hand printed Arabic characters; their technique is known as extracting the skeleton of the image by using the thinning algorithm [77]. Certain features are extracted, such as lines, curves and loops. This technique was tested by 10 users and achieved a 90.00% recognition rate. The system for printed Arabic character recognition proposed by Amin and Mansoor [78] consists of three steps. The first step is pre-processing, including such tasks as binarisation; the second step is extraction of global features such as number of dots and number of sub-words; finally, an ANN is used for classification. This system achieved a 98.00% recognition rate in testing on a few hundred words. Al-Badr and Haralick in [79] proposed a technique for recognition of machine-printed Arabic words without segmentation. This technique describes shape primitives as symbols; they are extracted by applying morphological operations on a word image and matched with symbol models. This technique was tested on 42,000 words and an accuracy of 73.00% on scanned text and 99.40% on noise free text was obtained. A system for recognition of multi-font Arabic cursive words without segmentation has been suggested by Khorsheed and Clocksin [54]. This technique extracts Fourier Transform coefficients from a normalised polar image. The words are represented by templates which include Fourier transform coefficients. Recognition is accomplished by calculating Euclidean distances from these templates. The recognition rate was over 90.00% but this technique failed with many fonts, especially handwritten words. Mahmoud, *et al.*, in [80] introduced an algorithm for Arabic character recognition, using the Modified Fourier Spectrum. Their algorithm estimated the Modified Fourier Spectrum descriptors of the contour of the primary shape of Arabic characters. Ten features were used in the training and testing. They reported 95.90% as the average recognition rate. A hybrid technique based on both NNs and HMMs for multifont Arabic characters recognition was proposed by Ben Amor and Ben Amara [56] using Hough transform for feature extraction. Their system was implemented on a set of 85,000 samples in five different fonts and the overall recognition rate was 97.36%. The system of Souici-meslati and Sellami in [81] uses a neuro-symbolic classifier for the recognition of handwritten Arabic literal amounts. They extracted structural features, then a symbolic knowledge base was constructed according to their features. The symbolic representation was converted into a NN to train on classified examples. Their system was applied to 1,200 words and achieved a 93.00% recognition rate. An approach for Persian text recognition was proposed by Jelodar, *et al.*, [82], using morphological operators to describe each sub-word and a template matching method for classification. This approach was tested on a set containing 3,000 words and achieved up to 99.90\% recognition rate. Al-Muhtaseb, *et al.*, in [62] introduced a system for recognition of printed Arabic texts based on HMMs. They extracted 16 features by using non-overlapping hierarchical windows. Their system evaluated the text with eight fonts and achieved recognition rates between of 98.08\% and 99.89%. A new technique using the C4.5 machine learning for printed Arabic text recognition has been proposed by Amin [63]. First, pre-processing tasks such as digitisation, connected component, and skew correction are applied. Then, global features such as number of sub-words, number of dots and their position are extracted. In classification, machine learning C4.5 is used for generating a decision tree for each word. This technique achieved a 92.00% recognition rate in testing on a set containing 1,000 printed Arabic words with different fonts. Khorsheed in [84] introduced a system for off-line recognition of omnifont Arabic texts. The system decomposes the text image into lines and a sliding window is used to extract statistical features. The features are fed to the HMM Toolkit (HTK) for recognition. The system was tested on data consisting of more than 600 A4-size sheets and achieved a 92.40% recognition rate.

There are algorithms which segment printed Arabic word into characters. Bushofa and Spann in [85] introduced a system of segmentation and recognition of printed Arabic characters. Their system started by segmenting the word into characters through finding the angle which occurs by the joining of two characters. The secondaries, such as dots and marks, are removed from the characters to reduce the number of classes and the details about the secondaries are stored for use in the classification stage. The structural features of the characters are extracted and fed to the final recognition stage. In the recognition stage, a decision tree is used to recognise the characters. This system was tested on a set of 4,260 samples and achieved a 97.23% recognition rate. Approaches for the automatic recognition of Arabic characters are introduced by Nawaz, *et al.*, [17]. Drift Correction is used for the problem of skewed images. The segmentation process consists of three steps. First, a horizontal projection is used to segment a document into lines; next, the line is segmented into words or sub-words by using the vertical projection method; finally, vertical projection is applied on the word level. A segmentation point is considered, if the value of the vertical projection is less than threshold. The features are extracted by using a centralised moments technique and passed to the ANN for classification. This approach yielded a recognition rate of approximately 76.00%. Zheng, *et al.*, in [40] presented a new algorithm for machine printed Arabic character segmentation. This algorithm is based on the vertical histogram as well as certain rules based on the characteristics of isolated characters. The segmentation process has three levels. Firstly, the horizontal histogram is used for line segmentation; secondly, the vertical histogram is employed for word and sub-word segmentation; finally, certain rules are used to test whether the sub-word includes only one character. Thereafter, the vertical histogram is applied on a sub-word which has more than one character to extract segmentation points. The algorithm was tested on more than 500 samples and achieved approximately 94.00% correct segmentation.

### 4.3. Handwritten Text Recognition

Handwritten text recognition entails greater challenges than printed text recognition owing to the fact its styles and sizes may vary for the same writer and different writers. Because of these difficulties, several researchers have conceived systems based on recognition of the whole word without segmentation. However, there are algorithms which split the words into characters.

**4.3.1. A Review of Handwritten Recognition without Segmentation**: A general framework of handwritten Arabic word recognition as a whole without segmentation consists of three stages: pre-processing, feature extraction and classification. The most significant researches on handwritten Arabic word recognition without segmentation are illustrated in this section. El-Hajj, *et al.*, in [11] has introduced a system for Arabic handwriting recognition without segmentation based on HMM. Their system uses baseline dependent features. The upper and lower baselines of words are extracted by using horizontal projection. The images are divided into vertical overlapping frames, 24 features being extracted from each frame. Two types of features are used: distribution features and concavity features. HMM is used for both training and testing stages. This system was evaluated on an IFN/ENIT database and achieved an average recognition rate of 86.51%. A two-stage system for Arabic handwriting recognition without segmentation was proposed by Al-Hajj, *et al.*, [86]. In the pre-processing, the upper and lower BLs of words is extracted. Features such as pixel density distribution and local pixel configurations are extracted from a binary image based on a sliding window method with different angles. Two stages are used for recognition. Three HMM-based classifiers are used in the first stage, while the second stage is based on the combination of these classifiers.

The system was tested on the IFN/ENIT database and achieved a recognition rate of up to 90.96%. In [87], Maddouri and Amiri proposed a system for handwritten Arabic word recognition. This system is based on combining a local and a global vision modelling for handwritten Arabic words. Structural features of the word, such as dots, loops, ascenders and descenders, are extracted as the global vision modelling of the word, while, normalised Fourier Descriptors are used to extract invariant statistical features of the word as a local vision modelling. A transparent NN is used, combining global and local vision modelling for recognition of handwritten Arabic words. This system was tested on 70 words of Arabic literal amounts and each word had 30 samples and achieved a 97.00% recognition rate. A new system for recognising Arabic handwritten words proposed by Al-Hajj, *et al.*, [88] relies on three HMM-based classifiers. The upper and lower baselines of words are extracted by using algorithm in Blumenstein, *et al.*, [89]. Vertical and slanted frames are used for extracting the features. Distribution features and concavity features are extracted from each frame [11]. Three homogeneous HMM classifiers are then combined and used at the decision level to increase the performance of the system. The same stages for combining the classifiers in Al-Hajj, *et al.*, [86] are used for recognition. This system was tested on using the IFN/ENIT database. Recognition rate achieved was higher than 90.00%. Dehghan, *et al.*, in [90] have put forward a holistic system for the recognition of Farsi/Arabic handwritten words based on HMM. The images are divided into fixed-width frames from right to left, the chain code histogram of the image a frame used as feature vectors and HMM is used for recognising the words. This system was tested on 17,000 images of 198 Iranian city names and achieved 65.00% recognition accuracy. A system for handwritten Arabic word recognition based on a semi-continuous one-dimensional HMM has been developed by Pechwitz and Maergner [91]. In the pre-processing stage, tasks such as normalisation, connected component analysis and extracting an image contour representation are performed. The normalised images are divided into frames and then features are extracted from each frame. The three frames are merged into one feature vector which is shifted from right to left. Up to 160 different HMM-models are used to represent the shapes of Arabic characters. This system was tested on the IFN/ENIT database and achieved 89.00% accuracy on word level. Benouareth, *et al.*, in [92] have applied a holistic approach to Arabic handwritten word recognition. The word image is divided into frames from right to left and 41 statistical and structural features being extracted from each frame. HMM is used to classify the word based on a likelihood criterion. This approach was tested on the IFN/ENIT database and achieved a 91.23% recognition rate. A system for classifying and retrieving Arabic handwritten text documents has been introduced by Brook and Al Aghbari [93]. This system is performed in three steps. Firstly, the text image is segmented into words, and then each word is segmented into sub-words. Secondly, structural and statistical features are extracted from these sub-words and then combined into feature vectors to represent a word. Finally, a NN is used for training and recognising the words. Ten different documents are used for classification: five documents for training and five documents for testing. This system achieved a 96.06% recognition rate. Xiang, *et al.*, in [94] have introduced a segmentation-free system for recognition of handwritten Arabic words based on HMM. This system uses a sliding window technique to divide the image into frames and each frame is divided into four cells. Distribution features and concavity features are then extracted from each frame, HMM being used for training and recognition. This system was tested on the IFN/ENIT database and achieved an average recognition rate of 84.09%. AIKhateeb, *et al.*, in [10] have introduced a system for word-based recognition of handwritten Arabic which is performed in three stages. Firstly, pre-processing tasks such as baseline estimation and normalisation are conducted, the normalisation algorithm in Pechwitz and Maergner [91] being used. Secondly, DCT is used to extract the features from the whole word. Finally, ANN is used for training and classification. This system used 500 words to experiment (80% for training and 20% for testing) extracted from the IFN/ENIT database and the recognition rate was 80%. The approach of Alma'adeed, *et al.*, in [61] for recognition of handwritten Arabic words is based on HMM. In the pre-processing, the word is normalised, its slant and slope are corrected and thinned and 29 features are then extracted from a normalised image, and HMM is used in the classification process. This system was tested on a database produced by 100 writers and achieved a recognition rate of about 45.00%. Another system for recognition of unconstrained Arabic handwritten words and based on multiple HMMs has been proposed by Alma'adeed, *et al.*, [95]. Similar pre-processing tasks to these proposed in Alma'adeed, *et al.*, [61] are performed. 29 features are extracted from the skeleton of the image. A global classifier and a local classifier are used to improve the accuracy in Alma'adeed, *et al.*, [61]. The global features such as ascenders, descenders, loops and dots are used to divide the words into groups. The local classifier uses multiple HMMs, where an HMM is used for each group to classify the groups into words. This system was tested on the same database as in Alma'adeed, *et al.*, [61] and achieved a variable recognition rate for each group of between 60.00% and 97.00%. Alma'adeed in [61] envisioned a system based on ANN for unconstrained Arabic handwritten word recognition. The global features such as a number of dots, loops ascenders and descenders are extracted from the normalised image of the whole words. ANN is used to classify the global features. The ANN has 3 layers (8 neurons as input layer, 40 neurons as a middle layer and 70 neurons as output layer) to classify 70 words. This system was tested on the database used in Alma'adeed, *et al.*, [61] and achieved a 63.00% recognition rate. AlKhateeb, *et al.*, in [96] have conceived a system of handwritten Arabic word recognition based on HMM. This system involves three steps. Firstly, pre-processing tasks include skew/slant correction and normalisation. Secondly, the word image is divided into frames and a set of features are extracted from each frame. Two types of features are used in this work. Intensity features are used for training and structure-like features are used for re-ranking for improved accuracy. Finally, these features are used for classification. This system was tested on the IFN/ENIT and achieved a recognition rate of 82.32% and 83.55% with re-ranking. A comparative study of performance of HMMs and Dynamic Bayesian Networks (DBNs) for classifying handwritten Arabic words has been conducted by AlKhateeb, *et al.*, [97]. Three stages are used in this work. First, this system uses pre-processing tasks in AlKhateeb, *et al.*, [98]. In the second stage, features based on a sliding window are extracted from each image. Finally, in the classification stage, HMMs and DBNs are used and compared. This system was tested on the IFN/ENIT database and the results revealed that HMMs achieved a higher recognition rate and lower complexity than DBNs. An approach for handwritten Arabic word recognition based on KNN has been described by AIKhateeb, *et al.*, [63]. This approach uses the pre-processing tasks of AlKhateeb, *et al.*, [98]. An image of a word is segmented into overlapping frames and the absolute mean value of all the coefficients in each frame computed as feature vectors of the word image. KNN is used to classify the words by using these features. This approach was tested on the IFN/ENIT database and achieved a recognition rate of 76.04%. Menasri, *et al.*, in [99] have introduced a system of Arabic handwriting recognition based on the shapes of the alphabet. They proposed a new shape-based alphabet to recognise handwritten Arabic words. Arabic characters are classified into root shapes and optional tails. They used a plain generic grapheme segmentation in Dupre [100] which was designed for Latin cursive writing. Baseline-dependant features described by Dupre [100] are extracted from the graphemes and a hybrid HMM/ANN is used for classification. HMM is used to represent each character-body class, while NN calculates the probability y distribution of observations. This system was tested on the IFN/ENIT database and achieved an 87.40% recognition rate. Another hybrid HMMs/ANN approach was presented by Zermi, *et al.*, [101]. Word images are segmented into strokes called graphemes, and global and local features are extracted from the graphemes. These graphemes are represented by a sequence of observations. The recognition process is based on a HMM which uses a feed-forward NN for classification. This system was tested on database containing 5,000 Arabic words (2,500 for training and 2,500 for testing) and achieved a recognition rate of approximately 95.00%.

Other investigators have assumed that the characters are already segmented, in order to avoid the segmentation process. These works have presented methods based on recognition of characters and digits. Awaidah and Mahmoud in [74] presented a system for recognition of handwritten Arabic (Indian) numerals based on HMM. Each image is divided into sub-regions. Features such as gradient, structure, and concavity are extracted from each sub-region. HMM is used for training and testing. This system was tested on database containing 21,120 digits written by 44 writers and it achieved a 99.00% recognition rate. The performance of several feature sets and classification techniques for classifying handwritten Arabic digits was evaluated by Abdleazeem and El-Sherif [102]. They proposed a new technique, local directional features, for extracting the features of handwritten Arabic digits. A strategy to select and design an optimal two-stage classification system for handwritten Arabic digits was proposed as well. They implemented their system on the ADBase [73] and achieved an average recognition rate of 98.67%. Asiri and Khorsheed in [59] have presented a system for recognition of handwritten Arabic forms based on ANN. Haar wavelet transform is used to extract features of the characters. Wavelet coefficients are used for training and classifying the characters using ANN. This system used 15,800 isolated characters written by 500 writers for experiment. They assessed the performance of the system via three different cases. For each case, the error rate was 26.00%, 18.00% and 12.00%. In [15], Mowlaei, *et al.*, have introduced a system for recognition of isolated handwritten Farsi/Arabic characters and numerals using Haar wavelet transform for feature extraction. Approximation coefficients at level three are extracted from the images and converted into feature vectors. The authors categorised 32 isolated characters in Farsi into 8 classes to classify them by ANN. This system was tested on a database containing 579 postal addresses in Iran and it yielded a recognition rate of 97.24%. Shanbehzadeh, *et al.*, in [103] have offered a new set of features for handwritten Farsi characters. These features are a combination of two parts to distinguish the characters. The first group of features describes the general structure of the character, such as the number of dots and their positions. In the second group, initially, the image of a character is divided into frames. Then, statistical information, such as distributive and concavity features, is extracted from each frame. These features are classified by Vector Quantization [104]. This system was tested on a database of 3,000 isolated characters and achieved 85.59% accuracy. Two types of features sets of handwritten Persian/Arabic digits recognition have been put forward by Alaei, *et al.*, [105]. One is based on chain-code directional frequencies of the image contour, while the other is based on transition features (horizontally and vertically). A multi-level support vector machine (SVM) is used for classification. This system was tested on 80,000 handwritten samples of Persian digits (60,000 for training and 20,000 for testing) and achieved 99.02% accuracy. Hamdi, *et al.*, in [106] have suggested a new system for handwritten Arabic isolated character recognition. This system uses Principal Component Analysis (PCA) and SVM classifier. The Moments and Fourier Descriptor of profile projection and centroid distance are used for feature extraction. This system was tested on database containing 1,000 handwritten Arabic isolated characters (800 used for training and 200 for testing) and achieved a 96.00% recognition rate. A system for recognition of handwritten Persian isolated characters based on HMMs has been proposed by Dehghani, *et al.*, [107]. Median and morphological filters are used to remove noise. Regional projection contour transformation is used to extract the features of the characters. In classification, HMMs were used and achieved a 71.82% recognition rate. Sabri in [108] has proposed a system based on HMM to recognise handwritten Arabic (Indian) numerals. In this system, features are based on the digit as a whole (not divided into frames). Features such as angle, distance, horizontal and vertical span are extracted from the digit image, HMM and Nearest Neighbour Classifier (NNC) are used and compared in the classification stage. This system was tested on database containing 21,120 samples written by 44 writers and achieved a 97.99% and a 94.35% recognition rate with HMM and NNC respectively.

**4.3.2. Handwritten Recognition based on Segmentation**: Off-line handwritten Arabic word recognition based on segmentation still represents a challenge in text recognition. Therefore, most researchers have presented methods based on recognition of the whole word without segmentation. However, there are algorithms which split the words into characters. The most significant researches segmenting a word into characters are illustrated in this section. Hamid and Haraty in [46] have developed a technique for segmenting handwritten Arabic texts. They scanned the whole text image and extracted a connected block of characters (BC). Topographic features of each BC, such as black pixel density, holes, and end points, are extracted to identify pre-segmentation points. ANN is used to verify these segmentation points. This algorithm was tested on 10,000 exemplars and achieved 69.72% accuracy. Saris, et al., in [45] have proposed an algorithm for Arabic character segmentation which is based on morphological rules of word contours. The contour of the word is extracted by using an algorithm in Pavlidis [109]. The extracting of morphological rules of the word is based on topological characteristics of the Arabic text. These rules are used to find segmentation points. This algorithm was tested on a small database containing 100 handwritten Arabic words and achieved 86.00% good segmentation. A new segmentation algorithm based on rational invariant segments features has been outlined by Abdulla, *et al.,* [110]. This algorithm starts with the evaluation of strokes or curved segments through the words. They used Freeman chain code algorithm to identify the coordinates of each pixel on the image contour to help in detecting the slopes between points. The directions of stroke are listed in two groups to prepare nominated segmentation points. Finally, segmentation points are obtained from the nominated points by applying certain conditions. This algorithm was tested on an AHD/AUST database containing 12,300 words and the IFN/ENIT database containing 26,400. The results were evaluated by manually observing the results of the algorithm and achieved segmentation rates of 95.66% and 90.58% respectively. In [39], Lorigo and Govindaraju have suggested an algorithm for segmentation of handwritten Arabic words based on over-segments. Firstly, they measured the stroke width obtained from image and detected connected components (CCs) for each word. Two techniques (gradient and down-up) are used to detect the candidate points. This algorithm was tested on 200 images from the IFN/ENIT database and reached a segmentation rate of 92.30%. Wshah, *et al.,* in [111] put forward an algorithm based on the contour and skeleton of the image for segmenting handwritten Arabic words. Firstly, preprocessing tasks, such as noise removal and smoothing, are applied on the image. The algorithm extracts intersection points from the skeleton of the image. The distance maps between intersection points and the corresponding paths on the image contour are computed to discover segmentation points. This algorithm was tested on over 6,300 words from DARPA and IFTN databases. Their results showed that 21.30% extracted exactly one character, 6.30% two characters and 3.30% three characters. The approach of Olivier, *et al.,* [112], however, is a segmentation algorithm based on segmenting the words into portions of characters called graphemes. They used the upper contour of a word to determine segmentation points and identified certain conditions to accept or reject segmentation points, such as no segmentation point in loops. This algorithm was tested on a database containing 6,000 words written by 20 writers and attained 98.52% accuracy. Motawa, *et al.*, in [47] have introduced an algorithm for segmenting handwritten Arabic words in which they assume that the characters are usually connected to subsequent characters by horizontal lines and the connection between characters is usually represented by strokes. These assumptions were used with mathematical morphological techniques to identify the segmentation points. This algorithm was tested on a database containing a few hundred words and achieved an 81.88% good segmentation rate. Al-Hamad and Abu Zitar in [113] have described a technique for segmenting a word into its primitives where an ANN is used for validating the segmentation points based on certain features. Firstly, an algorithm of Arabic Heuristic Segmentation (AHS) is used to segment a word into primitives. Then, the features of the structural characters are extracted by using the Modified Direction Features technique (MDF). Finally, these features are converted to an ANN for training and testing to validate the segmentation points. This system was tested on a database of 500 words written by 20 writers and achieved an 82.98\% good segmentation rate. Another type of segmentation proposed by AlKhateeb, *et al.*, [38] is word segmentation. In this technique, words are detected and extracted in a handwritten Arabic text. This method is based on distances between words and sub-words. The distances between the connected components are measured and analyzed to identify an optimal threshold for word segmentation. This technique was tested on 200 images collected from an IFN/ENIT database and achieved a correct segmentation rate of 85.00%.

Other researchers have proposed methods to recognise Arabic handwriting characters after segmentation. For example, Jannoud in [55] has described a system of handwritten Arabic text recognition which consists of several steps. Traditional preprocessing tasks, such as binarisation and noise removing were implemented to label each component in order to distinguish sub-words. The segmentation stage starts with finding horizontal gaps to segment a document into lines. Connected components are then extracted from each line. An algorithm is used to find the segmentation points through the baseline based on the cross points. Then, DWT is used to extract the features of each character. Finally, a minimum distance classifier is used to classify the characters. Various documents containing hundreds of characters are used for training and other, different, documents for testing. This system achieved a 90% recognition rate of characters. The system of Aljuaid, *et al.*, in [114] recognises Arabic handwritten texts using a genetic approach. The authors use horizontal and vertical projection of a thinned image. The image is segmented into small peaks based on the vertical projection. Structural features of each character, such as length, width and loop, are extracted to distinguish the shape of the character. Features of each peak are also extracted and fed to the genetic algorithm for classification. The system was tested on an AHPD-UTM database using 11,450 characters for training and 37,350 characters for testing. This system achieved an accuracy of approximately 87.00%. In [115], Lawgali, *et al.*, has presented a framework for the recognition of handwritten Arabic words. The framework proposed involves two phases (training phase and testing phase). In the training phase, the Arabic handwriting characters are used and highest value DCT coefficients of each character stored as features. HACDB database [75] containing 6,600 shapes of handwritten Arabic characters including overlapping characters were used in this phase. In the testing phase, handwritten Arabic words are segmented. Then DCT and ANN are used for feature extraction and classification, respectively. Classification is achieved in two steps (classification of the segmented characters and classification of the word). A dictionary is constructed and used to correct any error(s) occurring during the previous stages of the recognition process. This work has been tested with IFN/ENIT database [69] and achieved 90.73% the overall accuracy of correct classification

## 5. Discussion and Recommendations

In this paper, the problem of off-line recognition of handwritten Arabic words is addressed. There are two general approaches to off-line handwritten recognition. One is classification of the whole word without segmentation and the other is classification based on segmenting the word into characters and then classifying the word character by character [3]. The cursive nature of an Arabic script makes the segmentation of its words into individual characters a difficult task. Most of the currently proposed segmentation algorithms do not solve the problem of overlapping characters in Arabic handwriting, nor do they recognise the word after segmentation [2]. The segmentation stage is the most difficult stage and the main source of errors in recognition. Segmentation techniques still represent a challenge in text recognition and need to be improved [2]. Hence, researchers have put forward several methods based on the recognition of the whole word without segmentation, as well as other methods which assume that the characters have already been segmented, in order to avoid the segmentation process. However, recognition of whole words without segmentation has limitations because it can just classify the words trained; on the other hand, recognition based on segmentation does not have this restriction. Table 3 summarises the difference between recognition systems based on segmentation and those without segmentation.

**Table 3. Difference between Recognition Systems based on Segmentation and without Segmentation**

| | | |
|---|---|---|
| Application | Open up | Specific |
| Database | Large | Restricted |
| Challenges | More | Less |

There are a variety of applications that can be used when applying the segmentation based approach and with a large database for training characters, whereas the non-segmented approach is only useful with a restricted database owing to the retraining necessary for the classification of new words. However, the approach based on segmentation encompasses more challenges, one of which being the effect it has on the recognition rate, which is the main source of errors in recognition. In the non-segmented approach there is an absence of such challenges. Although promising results have been achieved, there are some suggestions for future work to improve this performance.

- A large database for handwritten Arabic characters containing all shapes of handwritten characters is needed. Digits (i.e. numerical numbers) are also needed to add to the database as some text might include them. This database may become a benchmark for handwritten Arabic characters and it will help researchers to compare their results.
- Arabic texts might have short marks. These marks, as shown in Figure 10, can be placed above or below characters. These marks are written as strokes and can affect recognition owing to their location either above or below the characters, like dots. Figure 11 shows marks and dots on an Arabic text. Therefore, these marks need to be recognised and a distinction made between them and the dots.

| Fathah | Kasrah | Dammah | Maddah | Sukun | Tanwin Fathah | Tanwin Kasrah | Tanwin Dammah | Shaddah | Hamza |
|---|---|---|---|---|---|---|---|---|---|
| ـَ | ـِ | ـُ | ـٓ | ـْ | ـً | ـٍ | ـٌ | ـّ | ء |

**Figure 10. Marks in Arabic Writing**

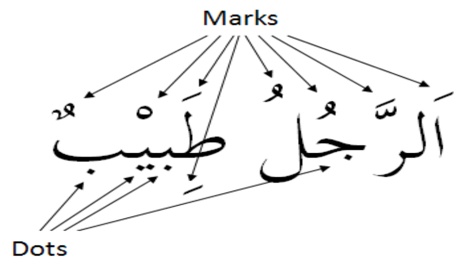



**Figure 11. Marks in Arabic Text**

- Segmentation is an important stage because it has an effect on the recognition performance. It still represents a challenge in text recognition since handwritten Arabic text can have different styles and sizes. Figure 12 shows some complex styles of handwritten Arabic words. Therefore, the segmentation algorithm should be improved by further investigating the more complex problem relating to style, and size.

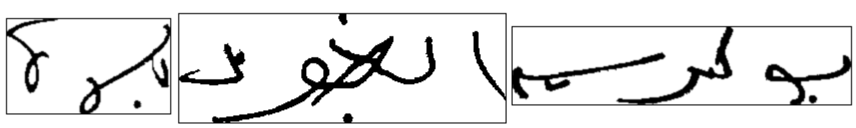


**Figure 12. Some Complex Styles of Handwritten Arabic Words**

- Due to the size variations of characters and the similarity between some of the handwritten characters, the structure of the features, such as loops, curves, and lines, can be extracted and classified into groups according to these features. This might improve the recognition performance.

## 6. Conclusion

This paper described a survey of Arabic character recognition systems. The most significant approaches to recognition of printed Arabic texts have been provided. However, such systems are not useful to deal with the inherent complexity of handwriting. The investigation into problems of recognition of handwritten Arabic words shows that most techniques developed so far are based on recognition of whole words without segmentation. This is due to the cursive nature and the peculiar ligatures of Arabic script. In addition, most of the segmentation algorithms currently proposed does not solve the problem of overlapping characters in Arabic handwriting, nor do they attempt to recognise the words after their segmentation. Some researchers have attempted to recognise the characters after the segmentation process but identifying characters only still does not provide a complete solution because this may lead to misclassifications of some characters in a word. Moreover, the objective of these systems is to identify the word, not a group of characters. Other researchers have assumed the characters are already segmented, in order to avoid the segmentation process and they have proposed techniques for isolated Arabic characters. However, to recognise characters after the segmentation process entails more challenges.

## References

[1]. P. Ahmed and Y. Al-Ohali, "Arabic Character Recognition: Progress and Challenges", Journal of King Saud University, **(2000)**, pp. 85-116.

[2]. A. M. Zeki, "The Segmentation Problem in Arabic Character Recognition the State of The Art", First International Conference on Information and Communication Technologies, **(2005)**, pp. 11-26.

[3]. L. M. Lorigo and V. Govindaraju, "Offline Arabic Handwriting Recognition", A Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, **(2006),** pp. 712-724.

[4]. Y. M. Alginahi, "A Survey on Arabic Character Segmentation", International Journal on Document Analysis and Recognition, **(2002)**, pp. 1-22.

[5]. A. M. Zeki and M. S. Zakaria, "Challenges in Recognizing Arabic Characters", The national conference for computer, Abu-Al-Aziz King University, Arabia Saudi, **(2004)**, pp. 445-453.

[6]. K. Jumari and M. Ali, "A Survey and Comparative Evaluation of Selected Off-Line Arabic Handwritten Character Recognition Systems", Jurnal Teknology, Malaysian University of Technology, **(2002)**, pp. 1-18.

[7]. R. Halavati, M. Jamzad and M. Soleymani, "A Novel Approach to Persion Online Hand writing Recognition", Transactions on Engineering, Computing and Technology, **(2005)**, pp. 232-236.

[8]. N. Mezghani, A. Mitiche and M. Cheriet, "On-Line Recognition of Handwritten Arabic Characters Using A Kohonen Neural Network", In Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition, **(2002)**, pp. 490-495.

[9]. R. Plamodon and S. N. Srihari, "Online and Offline Handwriting Recognition", A Comprehensive Survey, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 1, **(2000),** pp. 63-84.

[10]. J. H. AlKhateeb, R. Jinchang, J. Jianmin, S. S. Ipson and H. El-Abed, "Word-based Handwritten Arabic Scripts Recognition using DCT Features and Neural network Classifier", In 5th International Multi-Conference on Systems, Signals and Devices, **(2008)**, pp. 1–5.

[11]. R. El-Hajj, L. Likforman-Sulem and C. Mokbel, "Arabic Handwriting Recognition Using Baseline Dependant Features and Hidden Markov Modeling", In Proceedings of The Eighth International Conference on Document Analysis and Recognition, **(2005),** pp. 893–897, Washington, USA.

[12]. H. El-Abed and V. Margner, "Comparison of Different Preprocessing and Feature Extraction Methods for Offline Recognition of Handwritten Arabic Words", In Ninth International Conference on Document Analysis and Recognition, vol. 2, **(2007),** pp. 974-978.

[13]. H. Al-Rashaideh, "Preprocessing Phase for Arabic Word Handwritten Recognition", Russian Academy of Sciences, Russian Federation, vol. 6, no. 1, **(2006),** pp. 11-19.

[14]. F. Farooq, V. Govindaraju and M. Perrone, "Pre-processing Methods for Handwritten Arabic Documents", In Eighth International Conference on Document Analysis and Recognition, vol. 1, **(2005),** pp. 267–271.

[15]. A. Mowlaei, K. Faez and A. T. Haghighat, Feature Extraction with Wavelet Transform for Recognition of Isolated Handwritten Farsi/Arabic Characters and Numerals", In 14th International Conference on Digital Signal Processing, vol. 2, **(2002),** pp. 923–926.

[16]. A. Cheung, M. Bennamoun and N. W. Bergmann, "A Recognition-Based Arabic Optical Character Recognition System", In IEEE International Conference on Systems, Man, and Cybernetics, vol. 5, **(1998),** pp. 4189 –4194.

[17]. S. N. Nawaz, M. Sarfraz, A. Zidouri and W. G. Al-Khatib, "An Approach to Offline Arabic Character Recognition Using Neural Networks", In 10th IEEE International Conference on Electronics, Circuits and Systems, vol. 3, **(2003),** pp. 1328– 1331.

[18]. M. M. Altuwaijri and M. A. Bayoumi, "Arabic Text Recognition Using Neural Networks, In IEEE international Symposium on Circuits and Systems, vol. 6, **(1994),** pp. 415–418.

[19]. J. Alkhateeb, "Word Based Off-line Handwritten Arabic Classification and Recognition, PhD Thesis, School of Computing, Informatics and Media, University of Bradford, **(2010)**.

[20]. N. Arica and F. T. Yarman-Vural, "An Overview of Character Recognition Focused on Off-line Handwriting", IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 31, no. 2, **(2001),** pp. 216 –233.

[21]. R. Legault and C. Y. Suen, "Optimal Local Weighted Averaging Methods in Contour Smoothing", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, **(1997),** no. 8, pp. 801 –817.

[22]. W. L. Lee and K. C. Fan, "Document Image Pre-processing Based on Optimal Boolean Filter", Signal Processing, vol. 8, no. 1, **(2000),** pp. 45–55.

[23]. M. Khorsheed, “Off-line Arabic Character Recognition - A Review, Pattern Analysis Applications, vol. 5, **(2002),** pp. 31–45.
[24]. R. C. Gonzalez and R. E. Woods, “Digital Image Processing”, Dorling Kindersley Pvt. Ltd, **(2009)**.
[25]. J. Serra, “Morphological Filtering: An Overview”, Signal Processing, vol. 38, no. 1, **(1994),** pp. 3–11.
[26]. A. Atici and F. Yarman-Vural, “A Heuristic Algorithm for Optical Character Recognition of Arabic Scrip”, Signal Processing, vol. 62, no. 1, **(1997),** pp. 87–99.
[27]. J. Yang and X. Li, “Boundary Detection Using Mathematical Morphology”, Pattern Recognition Letters, vol. 16, no. 12, **(1995),** pp. 1277–1286.
[28]. A. AL-Shatnawi and K. Omar, “ A Comparative Study Between Methods of Arabic Baseline Detection”, In International Conference on Electrical Engineering and Informatics, vol. 1, **(2009),** pp. 73–77.
[29]. M. Pechwitz and V. Margner, “Baseline Estimation for Arabic Handwritten Words, In Eighth International Workshop on Frontiers in Handwriting Recognition, **(2002)**, pp. 479–484.
[30]. B. Parhami and A. Taraghi, “Automatic Recognition of Printed Farsi Texts, In Pattern Recognition, vol. 14, **(1981),** pp. 395–403.
[31]. P. Burrow, “Arabic Handwriting Recognition”, Master’s Thesis, University of Edinburgh, **(2004)**.
[32]. Z. Q. Liu, J. Cai and R. Buse, “Handwriting Recognition Soft Computing and Probabilistic Approaches”, Springer, **(2010)**.
[33]. B. A. Yanikoglu and P. A. Sandon, “Recognizing Off-line Cursive Handwriting”, In IEEE Computer Society Conference on Computer Vision and Pattern Recognition, **(1994)**, pp. 397 –403.
[34]. J. Dong, P. Dominique, A. Krzyyzak and C. V. Suen, “Cursive Word Skew/Slant Corrections Based on Radon Transform”, In Eighth International Conference on Document Analysis and Recognition, **(2005)**, pp. 478–483.
[35]. R. G. Casey and E. Lecolinet, “Strategies in Character Segmentation”, A Survey, In Proceedings of The Third International Conference on Document Analysis and Recognition, vol. 2, **(1995),** pp. 1028–1033.
[36]. Z. Al-Aghbari and S. Brook, “Hah Manuscripts: A Holistic Paradigm for Classifying and Retrieving Historical Arabic Handwritten Documents”, Expert Systems with Applications, vol. 36, no. 8, **(2009),** pp. 10942-10951.
[37]. G. Kim, V. Govindaraju and S. N. Srihari, “An Architecture for Handwritten Text Recognition Systems, International Journal on Document Analysis and Recognition, vol. 2, no. 1, **(1999),** pp. 37–44.
[38]. J. Alkhateeb, J. Ren, S. Ipson and J. Jiang, “Component-based Segmentation of Words from Handwritten Arabic Text”, International Journal of Computer Systems Science and Engineering, vol. 5, no. 1, **(2009),** pp. 54–58.
[39]. L. Lorigo and V. Govindaraju, “Segmentation and Pre-recognition of Arabic Handwriting”, In Proceedings of the Eighth International Conference on Document Analysis and Recognition, vol. 2, **(2002),** pp. 605-609.
[40]. L. Zheng, A. H. Hassin and X. Tang, “A New Algorithm for Machine Printed Arabic Character Segmentation”, Pattern Recognition Letters, vol. 25, **(2004),** pp. 1723–1729.
[41]. M. Altuwaijri and M. Bayoumi, “A New Thinning Algorithm for Arabic Characters Using Self-organizing Neural Network”, In IEEE International Symposium on Circuits and Systems, vol. 3, **(1995),** pp. 1824–1827.
[42]. M. Tellache, M. Sid-Ahmed and B. Abaza, “Thinning Algorithms for Arabic OCR”, In IEEE Pacific Rim Conference on Communications, Computers and Signal Processing, vol. 1, **(1993),** pp. 248 –251.
[43]. J. Cowell and F. Hussain, “Thinning Arabic Characters for Feature Extraction”, In Fifth International Conference on Information Visualisation, **(2001)**, pp. 181–185.
[44]. H. AL-Sadoun and A. Amin, “A New Structural Technique for Recognizing Printed Arabic Text”, International Journal of Pattern Recognition and Artificial Intelligence, vol. 9, no. 1, **(1995),** pp. 101–125.
[45]. T. Sari, L. Souici and M. Sellami, “Off-line Handwritten Arabic Character Segmentation Algorithm: ACSA”, In Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition, **(2002),** pp. 452–457.
[46]. A. Hamid and R. Haraty, “A Neuro-heuristic Approach for Segmenting Handwritten Arabic Text, In International Conference on Computer Systems and Applications, **(2001),** pp. 110-113.
[47]. D. Motawa, A. Amin and R. Sabourin, ‘Segmentation of Arabic Cursive Script”, In Proceedings of The 4th International Conference on Document Analysis and Recognition, **(1997),** pp. 625–628.
[48]. S. Al-Ma’adeed, “Recognition of Off-line Handwritten Arabic Words”, PhD Thesis, University of Nottingham, **(2004)**.
[49]. V. K. Govindan and A. P. Shivaprasad, “Character Recognition A Review”, Pattern Recognition, vol. 23, no. 7, **(1990),** pp. 671–683.
[50]. A. Amin and J. Mari, “Machine Recognition and Correction of Printed Arabic Text”, IEEE Transactions on Systems, Man and Cybernetics, vol. 19, no. 5, **(1989),** pp. 1300 –1306.
[51]. M. S. Khorsheed, “W. F. Structural Features of Cursive Arabic Script”, In Proceedings of The Tenth British Machine Vision Conference, **(1999)**, pp. 442 - 431.

[52]. G. N. Srinivasan and G. Shobha, "Statistical Texture Analysis", In Proceedings of World Academy of Science, Engineering and Technology, vol. 36, **(2008),** pp. 1264–1269.
[53]. K. M. Mohiuddin and J. Mao, "A Comparative Study of Different Classifiers for Handprinted Character Recognition, In Pattern Recognition in Practice IV, **(1994),** pp. 437–448.
[54]. M. S. Khorsheed and W. F. Clocksin, "Multi-font Arabic Word Recognition Using Spectral Features", In 15th International Conference on Pattern Recognition, vol. 4, **(2000),** pp. 543–546.
[55]. I. A. Jannoud, "Automatic Arabic Handwritten Text Recognition System", American Journal of Applied Sciences, vol. 4, **(2007),** pp. 857–864.
[56]. N. Ben Amor and N. E. Ben Amara, "Multifont Arabic Characters Recognition Using Hough Transform and HMM/ANN Classification", Journal of Multi-media, vol. 1, no. 2, **(2006),** pp. 50–54.
[57]. S. Alma'adeed, "Recognition of Off-line Handwritten Arabic Words Using Neural Network", In Geometric Modeling and Imaging-New Trends, **(2006)**, pp. 141–144.
[58]. K. Khatatneh, I. M. El-Emary and B. Al-Rifai, "Probabilistic Artificial Neural Network for Recognizing The Arabic Handwritten Characters", Journal of Computer Science, vol. 3, **(2006),** pp. 881–886.
[59]. A. Asiri and M. S. Khorsheed, "Automatic Processing of Handwritten Arabic Forms Using Neural Networks", In Proceedings Of World Academy of Science, Engineering And Technology, **(2005)**, pp. 147–151.
[60]. A. Dehghani, F. Shabani and P. Nava, "Off-line Recognition of Isolated Persian Handwritten Characters Using Multiple Hidden Markov Models", International Conference on Information Technology: Coding and Computing, **(2001)**, pp. 506–510.
[61]. S. Alma'adeed, C. Higgens and D. Elliman, "Recognition of Off-line Handwritten Arabic words using Hidden Markov Model Approach", In 16th International Conference on Pattern Recognition, vol. 3, **(2002),** pp. 481–484.
[62]. H. Al-Muhtaseb, M. Rami and S. Qahwaji, "Recognition of Off-line Printed Arabic Text Using Hidden Markov Models:, Signal Processing, vol. 88, **(2008),** pp. 2902–2912.
[63]. J. H. Alkhateeb, F. Khelifi, J. Jiani and S. S. Ipson, "A New Approach for Off-line Handwritten Arabic Word Recognition Using KNN Classifier", In IEEE International Conference on Signal and Image Processing Applications, **(2009)**, pp. 191–194.
[64]. S. Mozaffari, K. Faez, F. Faradji, M. Ziaratban and S. M. Golzan, "A Comprehensive Isolated Farsi/Arabic Character Database Handwritten OCR Research", 10th International Workshop on Frontiers in Handwriting Recognition, **(2006)**, pp. 385-389.
[65]. J. J. Hull, "A Database for Handwritten Text Recognition Research", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 16, **(1994),** pp. 550–554.
[66]. P. J. Grother, "Handprinted Forms and Character Database", Technical report, Technical Report and CDROM, National Institute of Standards and Technology, **(1995)**.
[67]. L. Deng, "The Mnist Database of Handwritten Digit Images for Machine Learning Research", Signal Processing Magazine, vol. 29, no. 6, **(2012),** pp. 141–142.
[68]. U. V. Marti and H. Bunke, "The Iam-Database: An English Sentence Database for Offline Handwriting Recognition", International Journal on Document Analysis and Recognition, vol. 5, no. 1, **(2002),** pp. 39–46.
[69]. M. Pechwitz, S. S. Maddouri, V. Mrgner, N. El- louze and H. Amiri, "IFN/ENIT - Database of Handwritten Arabic words", In Colloque Inter, Fran-cophone Sur lEcrit et le Document, CIFED2002, **(2002)**, pp. 129–136.
[70]. S. Al-Ma'adeed, D. Elliman and C. A. Higgins, "A Database for Arabic Handwritten Text Recognition Research, In Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition, **(2002)**, pp. 485 – 489.
[71]. S. Al-Ma'adeed, D. Elliman and C. A. Higgins, "A Database for Arabic Handwritten Text Recognition Research", The International Arab Journal of Information Technology, vol. 1, no. 1, **(2004),** pp. 117–121.
[72]. Y. Al-Ohali, M. Cheriet and C. Suen, "Databases for Recognition of Handwritten Arabic Cheques", Pattern Recognition, vol. 36, no. 1, **(2003),** pp. 111–121.
[73]. E. EL-Sherif and S. Abdleazeem, "A Two-Stage System for Arabic Handwritten Digit Recognition Tested on A New Large Database", In International Conference on Artificial Intelligence and Pattern Recognition, **(2007)**, pp. 237–242.
[74]. S. M. Awaidah and S. M., Mahmoud, "A Multiple Feature/Resolution Scheme to Arabic (Indian) Numerals Recognition Using Hidden Markov Models", Signal Processing, vol. 89, no. 6, **(2009),** pp. 1176-1184.
[75]. A. Lawgali, M. Angelova and A. Bouridane, "HACDB: Handwritten Arabic Characters Database for Automatic Character Recognition", In 4th European Workshop on Visual Information Processing, (**2013),** pp. 255–259.
[76]. A. Amin and H. Al-Sadoun, "Han Printed Arabic Character Recognition System", In Proceedings of The 12th International Conference on Pattern Recognition, vol. 2, **(1994),** pp. 536-539.
[77]. B. K. Jang and R. T. Chin, "One-pass Parallel Thinning: Analysis, Properties, and Quantitative Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 11, **(1992),** pp. 1129 –1140.

[78]. A. Amin and W. Mansoor, "Recognition of Printed Arabic Text Using Neural Networks", In Proceedings of The Fourth International Conference on Document Analysis and Recognition, vol. 2, **(1997),** pp. 612 –615.
[79]. B. Al-Badr and R. M. Haralick, "Segmentation-free Word Recognition with Application to Arabic", In Proceedings of the Third International Conference on Document Analysis aud Recognition, vol. 1, **(1995),** pp. 355 –359.
[80]. S. A. Mahmoud and A. S. Mahmoud, "Arabic Character Recognition Using Modified Fourier Spectrum", In Geometric Modeling and Imaging–New Trends, **(2006)**, pp. 155–159.
[81]. L. Souici-meslati and M. Sellami, "A hybrid Approach for Arabic Literal Amounts Recognition, The Arabian Journal for Science and Engineering, vol. 29, **(2004),** pp. 177–194.
[82]. M. S. Jelodar, M. J. Fadaeieslam, N. Mozayani and M. Fazeli, "A Persian OCR System Using Morphological Operators, In World Academy of Science, Engineering and Technology, **(2005)**, pp. 137–140.
[83]. A. Amin, "Recognition of Printed Arabic Text Based on Global Features and Decision Tree Learning Techniques", Pattern Recognition, vol. 33, **(2000),** pp. 1309–1323.
[84]. M. S. Khorsheed, "Offline Recognition of Omnifont Arabic Text Using The HMM Toolkit (HTK)", Pattern Recognition Letters, vol. 28, **(2007),** pp. 1563–1571.
[85]. B. M. F. Bushofa and M. Spann, "Segmentation and Recognition of Arabic Characters by Structural Classification", Image and Vision Computing, vol. 15, no. 3, **(1997),** pp. 167-179.
[86]. R. Al-Hajj, C. Mokbel and L. Likforman-Sulem, "Combination of HMM-based Classifiers for The Recognition of Arabic Handwritten Words", In Ninth International Conference on Document Analysis and Recognition, vol. 2, **(2007),** pp. 959–963.
[87]. S. S. Maddouri and H. Amiri, "Combination of Local and Global Vision Modelling for Arabic Handwritten Words Recognition", In Proceedings of The Eighth International Workshop on Frontiers in Handwriting Recognition, **(2002)**, pp. 128–135.
[88]. R. Al-Hajj, L. Likforman-Sulem and C. Mokbel, "Combining Slanted-Frame Classifiers for Improved HMM-Based Arabic Handwriting Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 7, **(2009),** pp. 1165-1177.
[89]. M. Blumenstein, C. K. Cheng and X. Y. Liu, "New Pre-Processing Techniques for Handwritten Word Recognition", In Proceedings of The 2nd IASTED Conference Visualization, Imaging and Image Processing, **(2002)**, pp. 480–484.
[90]. M. Dehghan, K. Faez, M. Ahmadi and M. Shridhar, "Handwritten Farsi (Arabic) Word Recognition: A Holistic Approach Using Discrete HMM", Pattern Recognition, vol. 34, no. 5, **(2001),** pp. 1057–1065.
[91]. M. Pechwitz and V. Maergner, " HMM Based Approach for Handwritten Arabic Word Recognition Using The IFN/ENIT – Database", In Proceedings of The Seventh International Conference on Document Analysis and Recognition, **(2003)**, p p . 890–894.
[92]. A. Benouareth, A. Ennaji and M. Sellami, "HMMs with Explicit State Duration Applied to Handwritten Arabic Word Recognition", In 18th International Conference on Pattern Recognition, vol. 2, **(2006),** pp. 897–900.
[93]. S. Brook and Z. Al-Aghbari, "Holistic Approach for Classifying and Retrieving Personal Arabic Handwritten Documents", In Proceedings of The 7th International Conference on Artificial Intelligence, knowledge Engineering and Databases, **(2008)**, pp. 565–570.
[94]. D. Xiang, H. Yan, X. Chen and Y. Cheng, "Offline Arabic Handwriting Recognition System Based on HMM", In 3rd IEEE International Conference on Computer Science and Information Technology, vol. 1, **(2010),** pp. 526 –529.
[95]. S. Alma'adeed, C. Higgins and D. Elliman, "Off-Line Recognition of Handwritten Arabic Words Using Multiple Hidden Markov Models", Knowledge-Based Systems, vol. 17, **(2004),** pp. 75–79.
[96]. J. H. Alkhateeb, J. Ren, J. Jiang and H. Al-Muhtaseb, "Offline Handwritten Arabic Cursive Text Recognition Using Hidden Markov Models and Re-ranking", Pattern Recognition Letters, vol. 32, **(2011),** pp. 1081–1088.
[97]. J. H. Alkhateeb, O. Pauplin, J. Ren and J. Jiang, "Performance of Hidden Markov Model and Dynamic Bayesian Network Classifiers on Handwritten Arabic Word Recognition", Knowledge-Based Systems, vol. 24, **(2011),** pp. 680-688.
[98]. J. Alkhateeb, J. Ren, S. S. Ipson and J. Jiang, "Knowledge based Baseline Detection and Optimal Thresholding for Words Segmentation in Efficient Pre-Processing of Handwritten Arabic Text", In 5th International Conference Information Technology: New Generation, **(2008)**, pp. 1158–1159.
[99]. F. Menasri, N. Vincent, E. Augustin and M. Cheriet, "Shape-Based Alphabet for Off-Line Arabic Handwriting Recognition", In 9th International Conference on Document Analysis and Recognition, vol. 2, **(2007),** pp. 969-973.
[100]. X. Dupre, "Reconnaissance De Lecriture Manuscrite", PhD Thesis, Uni.v Rene Descartes - Paris, **(2003)**.

[101]. N. Zermi, M. Ramdani and M. Bedda, "Arabic Handwriting Word Recognition Based on A Hybrid HMM/ANN Approach, International Journal of Soft Computing, vol. 2, **(2007),** pp. 5–10.
[102]. S. Abdleazeem and E. EL-sherif, "Arabic Handwritten Digit Recognition", International Journal on Document Analysis and Recognition, vol. 11, **(2008),** pp. 127–141.
[103]. J. Shanbehzadeh, H. Pezashki and A. Sarrafzadeh, "Features Extraction from Farsi Handwritten Letters", Proceedings of Image and Vision Computing, New Zealand, **(2007)**, pp. 35–40.
[104]. R. M. Gray, "Vector quantization", IEEE Acoust, Speech Signal Processing Magazine, **(1984)**, pp. 4–29.
[105]. A. Alaei, P. Nagabhushan and U. Pal, "Fine Classification of Unconstrained Handwritten Persian/Arabic Numerals by Removing Confusion Amongst Similar Classes", International Conference on Document Analysis and Recognition, **(2009)**, pp. 601–605.
[106]. R. Hamdi, F. Bouchareb and M. Bedda, "Handwritten Arabic Character Recognition Based on SVM Classifier", In 3rd International Conference on Information and Communication Technologies: From Theory to Applications, **(2008)**, pp. 1–4.
[107]. A. Dehghani, F. Shabini and P. Nava, "Off-Line Recognition of Isolated Persian Handwritten Characters Using Multiple Hidden Markov Models", In Proceedings International Conference on Information Technology: Coding and Computing, **(2001)**, pp. 506–510.
[108]. S. Mahmoud, "Recognition of Writer-independent Off-Line Handwritten Arabic (Indian) Numerals Using Hidden Markov Models", Signal Processing, vol. 88, no. 4, **(2008),** pp. 844-857.
[109]. T. Pavlidis, "Algorithms for Graphics and Image Processing", Murray Hill ed., New Jersey, **(1981)**.
[110]. S. Abdulla, A. Al-Nassiri and R. A. Salam, "Off-line Arabic Handwritten Word Segmentation Using Rotational Invariant Segments Features", The International Arab Journal of Information Technology, vol. 5, no. 2, **(2008),** pp. 200-208.
[111]. S. Wshah, S. Zhixin and V. Govindaraju, "Segmentation of Arabic Handwriting Based on Both Contour and Skeleton Segmentation", In 10th International Conference on Documents Analysis and Recognition, **(2009)**, pp. 793-797.
[112]. G. Olivier, H. Miled, K. Romeo and Y. Lecourtier, "Segmentation and Coding of Arabic Handwritten Words", In Proceedings of The 13th International Conference on Pattern Recognition, vol. 3, **(1996),** pp. 264 –268.
[113]. H. A. Al-Hamad and R. Abu Zitar, "Development of an Efficient Neural-Based Segmentation Technique for Arabic Handwriting Recognition", Pattern Recognition, vol. 43, no. 8, **(2010),** pp. 2773–2798.
[114]. H. Aljuaid, Z. Muhammad and M. Sarfraz, "A Tool to Develop Arabic Handwriting Recognition System Using Genetic Approach", Journal Computer Science, vol. 6, **(2010),** pp. 619–624.
[115]. A. Lawgali, M. Angelova and A. Bouridane, "A Framework for Arabic Handwritten Recognition Based on Segmentation", International Journal of Hybrid Information Technology, vol. 7, no. 5, **(2014),** pp. 413–428.