Off-line Handwritten Arabic Character Recognition: A Survey

Maad Shatnawi
Higher Colleges of Technology, Abu Dhabi, UAE

Abstract - *The automatic recognition of text on scanned images has several applications such as automatic postal mail sorting and searching in large volume of documents. Although Arabic handwritten text recognition has been addressed by many researchers, it remains a challenging task due to several factors. This paper presents an overview of off-line handwritten Arabic character recognition and summarizes the main technical challenges and characteristics of Arabic. It also investigates the relevant existing research carried out towards this perspective.*

Keywords: Handwritten character recognition, OCR, Arabic text recognition,

1. Introduction

Handwritten character recognition is the ability of computers to convert human writing into text either on-line or off-line. On-line recognition is performed by writing directly to a peripheral input device such as a personal digital assistant (PDA) and a cellular phone. Off-line recognition, also called optical character recognition (OCR), is the ability of computers to convert human writing into text by scanning. The early attempts in the area of Latin character recognition were made in the middle of the 1940s with the development of digital computers. One of the early attempts in Chinese character recognition was made in 1966. However, the first publication of Arabic text recognition was in 1975 [1].

Some research effort has been done on surveying handwritten Arabic character recognition. Amara and Bouslama [2] presented a review of the classification techniques used in the optical character recognition of Arabic script. Lorigo and Venu Govindaraju [3] reviewed off-line handwritten Arabic character recognition methods. Beg et al. [4] presented the current OCR products and reviewed the work done in the hardware domain of Arabic OCR. Harous and Elnagar [5] presented handwritten character-based parallel thinning Algorithms. AL-Shatnawi and Omar [6] reviewed the various Arabic baseline detection methods. Alginahi [7] surveyed Arabic character segmentation methods. Al-Shatnawi et al. [8] evaluated and compared skeleton Arabic character extraction methods. This paper reviews the main stages of handwritten Arabic OCR systems, presents the major technical challenges of these systems, and provides a comprehensive review of the research done in this field.

The remainder of this paper is organized as follows. The next section addresses the main technical challenges in the field of Arabic handwritten character recognition. Section 3 presents the main characteristics of Arabic writing. Section 4 briefly describes the five stages of Arabic character recognition. Arabic baseline detecting methods and approaches are presented in section 5. Section 6 focuses on the major classification techniques and approaches that are used in Arabic OCR. The use of synthetic data in Arabic handwritten OCR systems is illustrated in section 7, and concluding remarks are presented in Section 8.

2. Technical Challenges of Handwritten Arabic Character Recognition

Although the latest improvements in Arabic character recognition methods and systems are very promising, the automatic recognition of Arabic handwritten characters remains a challenging task due to many factors that are summarized as follows. The first factor is the lack of sufficient support in terms of funding, books, journals, and conferences at all levels including governments and research institutions, and the lack of interaction between researchers of this field [1]. Secondly, the lack of enough Arabic digital dictionaries and programming tools, and the absence of large public databases of Arabic handwritten characters and words when compared to English where large databases such as CEDAR [9] have been publicly available for a long time. The lack of large Arabic databases is, in part, due to the difficult, time consuming, and error prone process of generating ground truth1 for Arabic on the character level [10, 11]. Thirdly, the start of Arabic character recognition is very late compared to other languages such as Latin and Chinese. In addition to that, the unique characteristics of Arabic writing can be considered as a great challenge. The next section presents some of these characteristics.

3. Main Characteristics of Arabic Writing

Arabic is a native language for more than 250 million people. It is the third largest international language used by over one billion Muslims in their different religious activities. In addition to the Arabic language, there are several languages that use the Arabic alphabet, such as Urdu, Farsi (Persian), Pashto, Jawi, and Kurdish. The

[1] Ground truth refers to assigning correct symbolic text to images which will be used for learning.

Arabic text is written from right to left and is always cursive in both machine printed and handwritten text [12].

The Arabic alphabet set is composed of 28 basic letters which consist of strokes and dots. Dots, above and below the characters, play a major role in distinguishing some characters that differ only by the number or location of dots e.g. Ba (ب), Ta (ت), and Noon (ن).

Table 1. Different shapes of Arabic letters.

Character	Isolated	Beginning	Middle	End
Alef	ا			ـا
Ba	ب	بـ	ـبـ	ـب
Ta	ت	تـ	ـتـ	ـت
Tha	ث	ثـ	ـثـ	ـث
Jeem	ج	جـ	ـجـ	ـج
Ha'	ح	حـ	ـحـ	ـح
Kha	خ	خـ	ـخـ	ـخ
Dal	د			ـد
Thal	ذ			ـذ
Ra	ر			ـر
Zy	ز			ـز
Seen	س	سـ	ـسـ	ـس
Sheen	ش	شـ	ـشـ	ـش
Sad	ص	صـ	ـصـ	ـص
Dhad	ض	ضـ	ـضـ	ـض
Tah	ط	طـ	ـطـ	ـط
Dha	ظ	ظـ	ـظـ	ـظ
Ain	ع	عـ	ـعـ	ـع
Ghain	غ	غـ	ـغـ	ـغ
Fa	ف	فـ	ـفـ	ـف
Qaf	ق	قـ	ـقـ	ـق
Kaf	ك	كـ	ـكـ	ـك
Lam	ل	لـ	ـلـ	ـل
Meem	م	مـ	ـمـ	ـم
Noon	ن	نـ	ـنـ	ـن
Ha	ه	هـ	ـهـ	ـه
Waw	و			ـو
Ya	ي	يـ	ـيـ	ـي

Table 2. Numerals that are commonly used in Arabic writing.

Arabic	Indian
0	٠
1	١
2	٢
3	٣
4	٤
5	٥
6	٦
7	٧
8	٨
9	٩

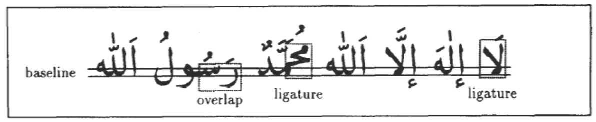

Figure 1. A sample of handwritten Arabic showing some of its characteristics [1].

The shape of an Arabic letter changes according to its location in the word, as shown in Table 1. For each character, there can be two to four different shapes: isolated, connected from the left (beginning of a word), connected from the left and right (middle of a word), and connected from the right (end of a word). Out of the 28 basic Arabic letters, six can be connected from the right side only while the other 22 can be connected from both sides. These six characters are: Alef (ا), Dal (د), Thal (ذ), Ra (ر), Zy (ز), and Waw (و) .These six characters have only two shapes, the isolated shape and the end shape, whereas the rest of the alphabets can appears in any of the four shapes mentioned above [13]. Consequently, each word may form one or more sub-words, where a sub-word is one or several connected characters, for example مستشفى, نور, and مناسبات. Moreover, in certain fonts, several characters can be vertically combined to form a ligature, especially in typeset and handwritten text. Ligatures can be formed out of two, three, or four characters [2]. Characters in a word may also vertically overlap without touching. Figure 1 presents some of these characteristics.

The use of special stress marks called diacritics is another distinguishing characteristic of Arabic. Diacritics such as Fat-ha (َ), Dhammah (ُ), Shaddah (ّ), Maddah (~), Sukun (ْ), and Kasrah (ِ) may change the pronunciation and the meaning of the word. The diacritics significantly affect the OCR performance [6, 14].

There are nine Arabic letters; Sad (ص), Dhad (ض), Tah (ط), Dha (ظ), Fa (ف), Qaf (ق), Meem (م), Ha (ه), and Waw (و) that have closed loops. This makes the closed loop an important feature in recognizing Arabic characters. One of the important characteristics of Arabic text is the presence of a baseline which is an imaginary horizontal line running through the connected portions of the text. If the script is handwritten, the baseline is not straight, and may only be estimated. Another feature of Arabic characters is that they do not have a fixed width or size, even in printed from. The character size varies according to its shape which is, in turn, a function of its position in the word.

In addition to the 28 characters, Arabic has additional non basic characters such as Hamzah (ء) and Ta marboota (ة). Hamzah can be isolated, on Alef (أ), on Waw (ؤ), or on Ya (ئ). Ta marboota is a special form of the letter Ta(ت) that only appears at the end of words. There are two types of numerals that are commonly used in Arabic; the Indian and Arabic numerals as shown in Figure 2.

4. Stages of Arabic Character Recognition

The process of Arabic character recognition consists of five stages; preprocessing, segmentation, feature extraction, classification, and post-processing [1]. The preprocessing enhances the raw images by reducing noise and distortion. This stage includes thinning, binarization, smoothing, alignment, normalization, and base-line detection.

Since Arabic text is cursive, segmentation is an important step in Arabic text recognition. Segmenting a page of text includes page decomposition and word segmentation. Page decomposition separates different logical parts, like text from graphics and lines of a paragraph, while word segmentation is the breakdown of words into isolated characters. The feature extraction stage analyzes a text segment and selects a set of structural or statistical features that can be used to uniquely identify the text segment. These features are extracted and passed in a form suitable for the recognition phase. Selecting the most suitable features plays a crucial role in the performance of the classification stage.

The recognition or classification stage is the main decision-making stage of an OCR system. This stage uses the features extracted in the previous stage to identify the text segments based on structural or statistical models. The classification stage uses machine learning techniques such as Artificial Neural Networks (ANN), support vector machines (SVM), k-nearest neighbors (*k*-NN), and Hidden Markov Models (HMM). The post-processing stage improves the recognition by refining the decisions taken by the previous stage and recognizes words by using context. It is ultimately responsible for outputting the best solution and is often implemented as a set of techniques that rely on character frequencies, lexicons, and other context information [1].

5. Arabic Baseline Detection Methods

Arabic baseline detecting methods can be categorized into four groups; the horizontal projection, the word skeleton, the word contour representation, and the Principal Components Analysis [6]. This section presents the current approaches of Arabic baseline detection. The horizontal projection method reduces the two-dimensional data into a one-dimension based on the pixels of the text image by summing up the pixel values of each row, and the row that obtains the highest score will be considered as the baseline. Although this method is easy to be implemented to printed text, it cannot easily detect handwritten text and it can be easily fooled by the diacritics.

Pechwitz and Maergner [15] used the word skeleton method to detect the baseline. This method creates the skeleton of the word based on polygon approximations. This method is applicable to both printed and handwritten Arabic text and is not affected by diacritics. However, it is computationally more expensive than other methods.

Farook et al. [16] detected the baseline according to the word Contour representation. This method finds the local minima of the word contour and then applies the linear regression technique to estimate the baseline of the text. This method can be applied to both printed and handwritten Arabic text and to both diacritized and non-diacritized text.

Burrow [17] applied the Principal Component Analysis for either foreground or background of the pixel distribution of the Arabic text in order to estimate the baseline direction and then applied the horizontal projection to detect the baseline. Detailed literature review of baseline detection methods can be found in [6].

6. Classification Approaches

There are several machine-learning (ML) classification techniques that are successfully applied to character recognition. These techniques can be classified into model-based and instance-based. ANN, HMM, and SVM are model-based classifiers that learn a model based on training examples to determine the decision boundaries between classes. Conversely, *k*-NN is an instance-based classification method, which assigns the class of the closet training examples in order to classify a new example. This section presents some of the existing approaches that use ML classification methods.

Graves and Schmidhuber [18] introduced a globally trained offline handwriting recognizer which is based on the raw pixel values of the input images. The system does not require any alphabet specific preprocessing and is applicable to any language. The two dimensional images were transformed into one dimensional label sequences of pixel values. The system employed the multidimensional long short-term memory (LSTM) neural networks as the classification technique. The system was evaluated on the IFN/ENIT database [10] where it outperformed the winner of ICDAR 2007 Arabic handwriting recognition contest [19] although neither author understands a word of Arabic.

Mahmoud and Awaida [20] described a technique for automatic recognition of off-line writer-independent handwritten Arabic (Indian) numerals using SVM and HMM. This work evaluated the use of the Gradient, Structural, and Concavity (GSC) features with a SVM classifier, and then the results are compared with HMM results using the same dataset and features. The SVM and HMM classifiers were trained with 75% of the data and tested with the remaining data. A two-stage exhaustive parameter estimation technique is used to estimate the best values for SVM parameters. The achieved average recognition rates were 99.83% and 99.00% using the SVM and HMM classifiers, respectively. The recognition rates of SVM proved to be superior to those of HMM for all digits and tested writers.

Natarajan et al. [21] evaluated HMM in handwritten character recognition. The authors concluded that the HMM-based system have several advantages because no pre-segmentation of words is required. On the other hand, this system suffers from two limitations; the assumption of conditional independence of the observations given the state sequence, and the restrictions on feature extraction imposed by frame-based observations. The use of pixel-level features from narrow slices of the text, specifically, the narrow windows provide very little contextual information making the conditional independence assumption in these systems unrealistic.

Hamdani et al. [22] presented an off-line handwriting recognition system based on the combination of multiple HMM classifiers. The classifiers are based on three on-line features and one off-line feature. The three on-line features are pixel values, densities and Moment Invariants, and pixel distribution and concavities. The authors used the technique described in [23] which allows having the on-line trace of the writing in a given image. The system was implemented using the HMM Toolkit (HTK) [24] and the IFN/ENIT database [10]. The authors concluded that the combination of on-line and off-line systems significantly improves the recognition accuracy.

Al-Hajj Mohamad et al. [25] proposed Arabic handwritten city names recognition based on combining three HMM classifiers that include a set of baseline-dependent and baseline-independent features. The three classifiers are combined at the decision level using three combination schemes: the sum rule, the majority vote rule, and an original neural network-based combination whose decision function is learned through candidate words' scores.

Mahmoud and Abu-Amara [26] describes a technique for the recognition of off-line handwritten Arabic (Indian) numerals using Radon and Fourier Transforms. A database of 44 writers with 48 examples of each digit totaling 21120 examples is used for training and testing the classifier. Radon-based features are extracted from Arabic numerals. Nearest Mean, *k*-NN, and HMM are used as digit classifiers. The recognition rates of these classifiers are 98.66%, 98.33%, 97.1%, respectively.

Parvez and Mahmoud [27] proposed a structural-based character recognition method. An Arabic text line is segmented into words/sub-words and dots are extracted. An adaptive slant correction algorithm that is able to correct the different slant angles of the different components of a text line is presented. A polygonal approximation algorithm is employed for text segmentation. Dynamic programming is used to select best hypotheses of a sequence of recognized characters for each word/sub-word. Prototype selection using set-medians, lexicon reduction using dot-descriptors are also utilized.

Tomeh et al. [28] and Habash and Roth [29] incorporated linguistically and semantically related features to Arabic character recognition systems. They presented an error detection system that uses deep lexical and morphological feature models to locate words or phrases that are likely incorrectly recognized. They used BBN's Byblos HMM-based off-line handwriting recognition system [30] to generate an *N*-best list of hypotheses for each segment of Arabic handwriting.

Sahlol et al. [31, 32] proposed a feed-forward back-propagation Neural Network approach. The method consists of four stages; binarization, normalization, noise removal, feature extraction, and classification. Three types of features were extracted; structural, statistical, and topological features. Structural features are the upper and lower profiles that capture the outlining shape of a connected part of the character as well as horizontal and vertical projection profiles. Statistical features include the four neighboring pixels for each pixel. Topological features include end points, pixel ration, and height to width ratio.

7. Use of Synthetic Data in Character Recognition

Large databases of Arabic handwritten characters and words are not publicly available when compared to Latin languages. Many papers in Arabic character recognition used their own small datasets such as Al -Badr and Mahmoud [1], Amin [33], and Khedher et al. [34], or they talked about large databases that are not available to public such as Kharma et al. [35], and Al-Ohali et al. [36]. Therefore, the use of synthetic data in building character recognition systems of different languages has been discussed and examined by many researchers.

Margner and Pechwitz [37] introduced a system for automatic generation of synthetic printed data for Arabic OCR systems. This system can be described as follows. First, the Arabic text has to be typeset. Then, a noise-free bitmap of the document and the corresponding GT is automatically generated. And finally, an image distortion can be superimposed on the character or word image to simulate the expected real world noise of the intended application.

Elarian et al. [38] presented an approach to synthesize Arabic handwriting text. First, real word images are segmented into labeled characters which are then concatenated in an arbitrary way to synthesize artificial word images. The nearest Euclidean-distance neighbor is used for matching characters that can be concatenated to produce natural-looking words. This synthesized text is used to train and test OCR system. Although the proposed approach is still infancy and was tested on only two writers from the IFN/ENIT dataset [10], the authors reported promising results.

Dinges et al. [39] presented a method for Arabic handwriting synthesis using Active Shape Models (ASM) computed based on 28046 online samples of multiple writers. ASMs were used to generate unique letter representations for each synthesis. Subsequently these representations were modified by affine transformations, smoothed by B-Spline interpolation and composed to text.

Shatnawi and Abdallah [40] extended the spatial congealing technique [41] and evaluated it on Arabic characters. Congealing models human distortions in the examples of one or more handwritten characters, and then uses this model to generate synthetic examples of other characters that are similarly distorted. The extended congealing approach along with different types of systematic affine transformations including rotation, scaling, and shearing were used to synthesize a large number of virtual training examples of isolated Arabic characters. Experimental results proved significant improvements across three machine-learning classification algorithms; k-NN, Naïve Bayes, and SVM.

8. Conclusion

This survey is focused on off-line handwritten Arabic character recognition. The most important challenges that face handwritten Arabic OCR systems and the main characteristics of Arabic writing were addressed. We investigated the major relevant existing approaches carried out towards this perspective. There are several approaches in this field. However, the research in handwritten Arabic character recognition is still in an early stage when compared to Latin and other languages.

Acknowledgment

The author would like to thank Prof. Boumediene Belkhouche for reviewing an earlier draft of this work.

References

[1] Al-Badr B, Mahmoud SA (1995) Survey and Bibliography of Arabic Optical Text Recognition. Elsevier Science B.V., Signal Processing 41: pp 49-77.

[2] Amara N, Bouslama F (2003) Classification of Arabic script using multiple sources of information: State of the art and perspectives. International Journal on Document Analysis and Recognition, 5(4): 195-212.

[3] Lorigo LM, Govindaraju, V (2006) Off-line Arabic Handwritten Recognition: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. EEE Trans. Pattern Analysis and Machine Intelligence 28(5): 712–724.

[4] Beg A, Ahmed F, Campbell P (2010) Hybrid OCR Techniques for Cursive Script Languages –A Review and Applications. 2nd International Conference on Computational Intelligence, Communication Systems and Networks, Liverpool, United Kingdom.

[5] Harous S, Elnagar A (2009) Handwritten Character-Based Parallel Thinning Algorithms: A Comparative Study. University of Sharjah Journal of Pure & Applied Sciences 6(1): 81-101.

[6] AL-Shatnawi A, Omar K (2008) Methods of Arabic Language Baseline Detection – The State of Art, IJCSNS International Journal of Computer Science and Network Security, 8(10).

[7] Alginahi, Y. M. (2013). A survey on Arabic character segmentation. International Journal on Document Analysis and Recognition (IJDAR), 16(2), 105-126.

[8] AL-Shatnawi, A. M., AlFawwaz, B. M., Omar, K., Zeki, A. M. (2014). Skeleton extraction: Comparison of five methods on the Arabic IFN/ENIT database. In Computer Science and Information Technology (CSIT), 2014 6th International Conference on (pp. 50-59). IEEE.

[9] CEDAR (Center of Excellence for Document Analysis and Recognition) (2006) USPS Office of Advanced Technology Database of Handwritten Cities, States, ZIP Codes, Digits, and Alphabetic Characters, [online]. Available at http://www.cedar.buffalo.edu.

[10] Pechwitz M, Maddouri S, Margner V, Ellouze N, Amiri H (2002) IFN/ENIT - Database of Handwritten Arabic Words. CIFED: 1-8.

[11] Kanungo T, Resnik P, Mao S, Kim D, Zheng Q (2005) The Bible and Optical Character Recognition. Communications of the ACM, 48(6): 124-130.

[12] Khorsheed M (2002) Off-Line Arabic Character Recognition – A Review. Pattern Analysis and Applications, 5:31-45.

[13] Al-Shoshan AI (2006) Arabic OCR Based on Image Invariants. Proceedings of the Geometric Modeling and Imaging ― New Trends (GMAI'06), IEEE.

[14] Zeki AM (2005) The segmentation problem on Arabic character recognition – the state of the art. 1st International Conference on Information and Communication Technology (ICICT) Karachi, Pakistan: 11-26.

[15] Pechwitz M, Maergner V (2002) Baseline estimation for Arabic handwritten words. Frontiers in Handwritin Recognition: 479–484.

[16] Farooq F, Govindaraju V, Perrone M (2005) Preprocessing Methods for Handwritten Arabic Documents. (ICDAR'05) Proceedings of the 2005

Eight International Conference on Document Analysis and Recognition, IEEE 1: 267-271.

[17] Burrow P (2004). Arabic handwriting recognition. [M.Sc. thesis]. Edinburgh (England): University of Edinburgh.

[18] Graves A, Schmidhuber J (2008) Offline handwriting recognition with multidimensional recurrent neural networks. 22sd Conference on Neural Information Processing Systems (NIPS).

[19] Margner V, Abed HE (2007) Arabic handwriting recognition competition. 9th International Conference on Document Analysis and Recognition (ICDAR 2007) 2: 1274–1278, Washington, DC, USA, IEEE Computer Society.

[20] Mahmoud SA, Awaida SM (2009) Recognition of Off-line Handwritten Arabic (Indian) Numerals Using Multi-scale Features and Support Vector Machines Vs. Hidden Markov Models. The Arabian Journal for Science and Engineering (AJSE), 34: 429-444.

[21] Natarajan P, Subramanian K, Bhardwaj A, Prasad R (2009) Stochastic Segment Modeling for Offline Handwriting Recognition. 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, Spain.

[22] Hamdani M, Abed HE, Kherallah M, Alimi AM (2009) Combining Multiple HMMs Using On-line and Off-line Features for Off-line Arabic Handwritten Recognition. 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, Spain.

[23] Elbaati A, Kherallah M, Ennaji A, Alimi AM (2009) Temporal Order Recovery of the Scanned Handwriting. 10th International Conference on Document Analysis and Recognition (ICDAR 2009), Barcelona, Spain.

[24] Young S, Evermann G, Gales M, Hain T, Kershaw D, Liu XA, Moore G, Odell J, Ollason D, Povey D, Valtchev V, Woodland P (2006) The HTK Book. Cambridge University, England.

[25] Al-Hajj Mohamad, R., Likforman-Sulem, L., and Mokbel, C. (2009). Combining slanted-frame classifiers for improved HMM-based Arabic handwriting recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(7), 1165-1177.

[26] Mahmoud SA, Abu-Amara MH (2010) Recognition of Handwritten Arabic (Indian) Numerals using Radon-Fourier-based Features. 9th WSEAS international conference on Signal processing, robotics and automation (ISPRA '10), Cambridge, UK, portal.acm.org.

[27] Parvez, M. T., and Mahmoud, S. A. (2013). Arabic handwriting recognition using structural and syntactic pattern attributes. Pattern Recognition, 46(1), 141-154.

[28] Tomeh, N., Habash, N., Roth, R., Farra, N., Dasigi, P., and Diab, M. T. (2013). Reranking with Linguistic and Semantic Features for Arabic Optical Character Recognition. In ACL (2): 549-555.

[29] Habash, N., and Roth, R. M. (2011, June). Using deep morphology to improve automatic error detection in Arabic handwriting recognition. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1: 875-884.

[30] Saleem, S., Cao, H., Subramanian, K., Kamali, M., Prasad, R., and Natarajan, P. (2009) Improvements in BBN's HMM-based offline Arabic handwriting recognition system. In 10th International Conference on Document Analysis and Recognition (ICDAR'09): pp. 773-777. IEEE.

[31] Sahlol, Ahmed T., Cheng Y. Suen, Mohammed R. Elbasyouni, and Abdelhay A. Sallam. (2014) A Proposed OCR Algorithm for the Recognition of Handwritten Arabic Characters. Journal of Pattern Recognition and Intelligent Systems.

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

[33] Amin A (1998) Off-line Arabic Character Recognition: The State of the Art. Pattern Recognition Society, 31(5): 517-530.

[34] Khedher M, Abandah G, Al-Khawaldeh A (2005) Optimizing Feature Selection for Recognizing Handwritten Arabic Characters. Proceedings of the Second World Enformatika Conference, WEC'05, Istanbul, Turkey: 81-84.

[35] Kharma N, Ahmed M, Ward R (1999) A New Comprehensive Database of Hand-written Arabic Words, Numbers and Signatures used for OCR Testing. IEEE Canadian Conference on Electrical and Computer Engineering, Alberta, Canada: 766-768.

[36] Al-Ohali Y, Cheriet M, Suen C (2000) Database For Recognition of Handwritten Arabic Cheques. Proceedings of the Seventh International Workshop on Frontiers in Handwriting Recognition, Amsterdam: 601-606.

[37] Margner V, Pechwitz M (2001) Synthetic Data for Arabic OCR System Development. In: Sixth International Conference on Document Analysis and Recognition (ICDAR'01), IEEE: 1159-1163.

[38] Elarian YS, Al-Muhtaseb HA, Ghouti LM (2011) Arabic Handwriting Synthesis. 1st International Workshop on Frontiers in Arabic Handwriting Recognition.

[39] Dinges L, Al-Hamadi A, and Elzobi M (2013) An Approach for Arabic Handwriting Synthesis Based on Active Shape Models. In Document Analysis and

Recognition (ICDAR), 2013 12th International Conference on. IEEE, 1260–1264.

[40] [40] Shatnawi M. and Abdallah S. (2015). Improving Handwritten Arabic Character Recognition by Modeling Human Handwriting Distortions. ACM Transactions on Asian and Low-Resources Language Information Processing.

[41] [41] Learned-Miller E (2006) Data Driven Image Models through Continuous Joint Alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(2) pp 236-250.