International Journal of Civil Engineering and Technology (IJCIET)
Volume 9, Issue 13, December 2018, pp. 207-213, Article ID: IJCIET_09_13_023
Available online at http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=9&IType=13
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

SEQUENTIAL SEGMENTATION FOR ARABIC HANDWRITING USING SIMILARITY MEASURES

Ahmed Kawther Hussein
Department of Computer Science, College of Education, Mustansiriyah University, Baghdad-Iraq

ABSTRACT

Arabic handwriting recognition is an interesting pattern recognition research area. Various researchers have tackled this problem from various perspectives. Some of them have used neural networks or complex classifiers approaches based on huge computational features while others have developed heuristic based approaches using the style of writing Arabic letters. In this article, a sequential segmentation for Arabic handwriting recognition matching was developed. The approach starts with testing set of segmentation widths in order to do the best segmentation of Arabic letters from the word. It matches each potential character with a stored dataset of Arabic letters. Next, the letter that achieves the best matching score is selected to be the result. Two types of matching techniques are used, the first one is point to point matching and the second one is correlation matching. Testing this approach on set of Arabic words, we observed that the approach has performed good segmentation results with ability of detecting the letters in various positions in the Arabic word.

Keywords: Arabic Handwriting Recognition, Sequential Segmentation, Pattern Recognition

Cite this Article: Ahmed Kawther Hussein, Sequential Segmentation for Arabic Handwriting Using Similarity Measures, International Journal of Civil Engineering and Technology, 9(13), 2018, pp. 207–213

http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=9&IType=13

1. INTRODUCTION

Arabic handwriting recognition is an important research area [8], [2], [4]. It has wide range of challenges, some of them are due to the way of writing Arabic letters that follows a curve pattern with high degree of freedom, and others are due to the intersection between the different letters in one word [8]. His literature contains various approach for Arabic handwriting recognition [2], [1] [8] have developed an Arabic handwriting recognition system without explicit segmentation using hidden Markov models (HMMs). They extracted statistical and

geometrical features. The classifier is combined of training and recognition model. The role of the training model is to estimate the parameters of the HMMs using Baum-Welchalgorithm. The recognition phase performs an extraction of features from the image and use character models. The Viterbi algorithm is used to determine the best output hypotheses. Benchmark IFN/ENIT database has been used to show the improvement of the model. Recurrent neural networks (RNNs) with Long Short-Term memory cells was used by [7]. Dropout also is used with maintaining the recurrent connections. Some researchers have used hybrid NN/HMM model for online Arabic handwriting recognition as [9]. Similarly, Hidden Markov Models (HMMs) were used. In addition to that, Multi-Layer Perceptron Neural Networks (MLPNNs) was used. Continuous strokes called segments were used as segmentation of the input signal using Beta- Elliptical strategy through searching for the extremum points of the curvilinear velocity profile.

- In the work of [5] Arabic handwriting based on DCT and HOG features is proposed and in the work of [6] simple and novel methodology to authenticate Arabic handwriting characters. Reaching our aim, we built our own character database. The research methodology depends on two stages: The first is character extraction where preprocessing the word and then apply segmentation process to obtain the character. The second is the character recognition by matching the characters comprising the word with the letters in the database.
- [1] Have developed an Arabic handwriting synthesis system with two concatenation models to synthesize Arabic words from segmented characters. They adopted Extended-Glyphs connection and Synthetic-Extensions connection. The authors use the system to synthesize handwriting from a collected dataset and inject it into an expanded dataset.
- In the work of [3] context dependent modeling for Arabic handwriting recognition was improved. CART trees are used for state tying because the number of parameters in context dependent models is high. New set of questions for the CART tree construction as used for categorization of the Arabic shapes based on “lossy mapping”. A combination of Hidden Markov Models and Recurrent Neural Networks using the hybrid approach was used.
- It can be seen from the literature review; various researchers have tackled this problem from various perspectives. Some of them have used neural networks or complex classifiers approaches based on huge computational features while others have developed heuristic based approaches using the style of writing Arabic letters. In this article a simplified model for Arabic handwriting recognition is presented. The remaining of the article is organized as follows. In section 2, we provide the methodology. While in section 3, we give the experimental results and discussion. The summary and future work is provided in section 4.

2. METHODOLOGY

2.1. a- Problem Definition

Characters segmentation is an important phase for both hand and printed writing segmentation. Each language has its own characteristics of writing which makes the segmentation process easier or more challenging according to the nature of characters and its way of writing. Arabic characters have a curving nature and they are written in connected approach. This makes the segmentation process more challenging. In this article, a matching approach based on stored dataset of the Arabic letters and the subject text is performed. The letter that achieves the

highest similarity is considered as the recognized letter. Various measures can be used for determining the matched letter.

2.2. Dataset description

We used a dataset combined of 32 templates of Arabic letters, which each template has four cases isolated, initial, medial, and final. Each template letters are represented as JPEG images with adequate resolution to preserve the letter writing. The template database is shown in Table 1.

Table 1 Arabic letters in different parts of the words –database

	Initial	Medial	Final	Isolated
1	ا	ـا	ـا	ا
2	بـ	ـبـ	ـب	ب
3	تـ	ـتـ	ـت	ت
4	ثـ	ـثـ	ـث	ث
5	جـ	ـجـ	ـج	ج
6	حـ	ـحـ	ـح	ح
7	خـ	ـخـ	ـخ	خ
8	د	ـد	ـد	د
9	ذ	ـذ	ـذ	ذ
10	ر	ـر	ـر	ر
11	ز	ـز	ـز	ز
12	سـ	ـسـ	ـس	س
13	شـ	ـشـ	ـش	ش
14	صـ	ـصـ	ـص	ص
15	ضـ	ـضـ	ـض	ض
16	طـ	ـطـ	ـط	ط

2.3. b- Algorithm

The algorithm of segmenting Arabic letters from an image that contains set of words is presented in the Table2. The input of the algorithm is an image with one line of Arabic writing. The output of the Table 2 pseudo code for segmenting Arabic letters based on similarity measures and stored templates.

Table 2 algorithm of segmenting Arabic letters from an image

Input
Input image

Output
Sentence

Starts:
WordsImages=extractWordImages(Sentence)
sentence=[]
For each wordImage in WordsImages

```
scaledWordImage=Rescale(wordImage,scalingSize)
    while ContainLetters(scaledWordImage)
[segmentedLetter,width]=segment(scaledWordImage)
          sentence=[sentence segmentedLetter]
scaledWordImage =cropImage(scaledWordImage, width)
      end
end


function [segmentedLetter,width]=segment(scaledWordImage)
similarity=0;
segmentedLetter=[]
for each of width segmentation Widths
segmentedLetterNew=Segment(scaledWordImage)
similarityNew=performMatching(Dataset, segmentedLetter)
                    if(similarityNew>similarity)
                    similarity= similarityNew
segmentedLetter=segmentedLetterNew
segmentedWidth=width
                    end
           end
end



End
```

3. EXPERIMENYAL RESUULTS

In order to evaluate our approach, table -3- shows the difference between the performance of our approach that has used the correlation measure and case of using the point to point matching as the case of the benchmark (M. A. Abdullah et al., 2012). It is observed that our approach has produced more accurate segmentation for most of the letters comparing with the benchmark.

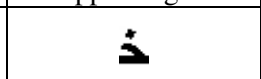
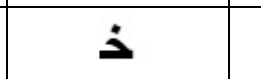
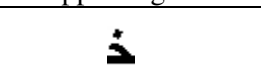
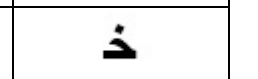
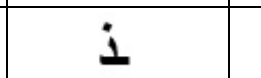
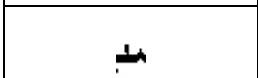
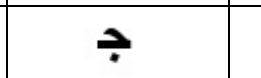
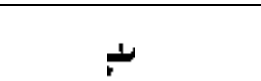
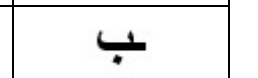
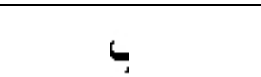
The Tables 3 show the results of the cropped and matched letters based on various Arabic words.

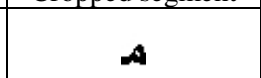
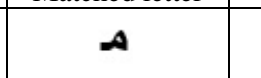
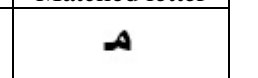
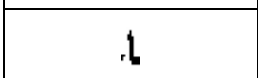
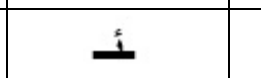
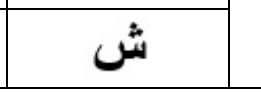
Table 3 comparison between benchmark and our approach

word	benchmark		Our approach	
	Cropped segment	Matched letter	Cropped segment	Matched letter
عشب				
				
				
				

word	benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
زرع				
				
				
				

word	benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
جامعة				
				
				

word	benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
خشب				
				
				
				

word	benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
ماء				
				
				

word	benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
زيت				
				
				
				

word	Benchmark		Correlation coefficient	
	Cropped segment	Matched letter	Cropped segment	Matched letter
هنا				

4. CONCLUSION

In this work, a sequential segmentation for Arabic handwriting recognition matching was developed. The approach starts with testing set of segmentation widths in order to do the best segmentation of Arabic letters from the word. It matches each potential character with a stored dataset of Arabic letters.

Next, the letter that achieves the best matching score is selected to be the result. Two types of matching techniques are used, the first one is point to point matching and the second one is correlation matching. Testing this approach on set of Arabic words, we observed that the approach has performed good segmentation results with ability of detecting the letters in various positions in the Arabic word.

A future work is to validate the developed approach on various similarity measures and to incorporate set of trained classifiers.

ACKNOWLEDGMENTS

The author would like to thank the Mustansiriyah university (www.uomustansiriyah.edu.iq) Baghdad - Iraq for its support in the present work.

REFERENCES

[1] Elarian, Yousef, Irfan Ahmad, Sameh Awaida, Wasfi G. Al-Khatib, and Abdelmalek Zidouri. "An Arabic handwriting synthesis system." Pattern Recognition 48, no. 3 (2015): 849-861.

[2] Ghaleb, Hashem, P. Nagabhushan, and Umapada Pal. "Segmentation of offline handwritten Arabic text." In Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop on, pp. 41-45. IEEE, 2017.

[3] Hamdani, Mahdi, Patrick Doetsch, and Hermann Ney. "Improvement of context dependent modeling for Arabic handwriting recognition." In Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on, pp. 494-499. IEEE, 2014.

[4] Hassen, H., & Al-Maadeed, S. Arabic handwriting recognition using sequential minimal optimization. Arabic Script Analysis and Recognition (ASAR), 2017 1st International Workshop On, 2017, 79–84.

[5] Kadhm, Mustafa S., and Alia Karim Abdul Hassan. "Arabic Handwriting Text Recognition Based on Efficient Segmentation, DCT and HOG Features." International Journal of Multimedia and Ubiquitous Engineering 11, no. 10 (2016): 83-92.

[6] Abdullah, Manal A., Lulwah M. Al-Harigy, and Hanadi H. Al-Fraidi. "Off-line arabic handwriting character recognition using word segmentation." arXiv preprint arXiv: 1206.1518 (2012).

[7] Pham, Vu, Théodore Bluche, Christopher Kermorvant, and Jérôme Louradour. "Dropout improves recurrent neural networks for handwriting recognition." In Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on, pp. 285-290. IEEE, 2014.

[8] Rabi, Mouhcine, Mustapha Amrouch, and Zouhair Mahani. "Cursive Arabic Handwriting Recognition System Without Explicit Segmentation Based on Hidden Markov Models." Journal of Data Mining and Digital Humanities (2018).

[9] Tagougui, Najiba, Houcine Boubaker, Monji Kherallah, and Adel M. Alimi. "A hybrid NN/HMM modeling technique for online Arabic handwriting recognition." arXiv preprint arXiv: 1401.0486 (2014).