International Journal of Computer Applications Technology and Research
Volume 4– Issue 12, 928 - 932, 2015, ISSN: 2319–8656

Isolated Arabic Handwritten Character Recognition Using Linear Correlation

Abd Elhafeez Hamid
Dept. of CS
Faculty of CS & IT
Kassala University
Kassala-sudan

Mariam M. Musa
Dept. of CS
Faculty of CS & IT
Kassala University
Kassala-sudan

Mona M. Osman
Dept. of CS
Faculty of CS & IT
Kassala University
Kassala-sudan

Moahib M. Alamien
Dept. of CS
Faculty of CS & IT
Kassala University
Kassala-sudan

Marwa M. Elsaied
Dept. of CS
Faculty of CS & IT
Kassala University
Kassala-sudan

Abstract: Handwriting recognition systems have emerged and evolved significantly, especially in English language, but for the Arabic language, such systems did not find that sufficient attention in comparison to other languages .Therefore, the aim of this paper to highlight the Optical Character Recognition using linear correlation algorithm in two dimensions and then the programs can to identify discrete Arabic letters application started manually, the program has been successfully applied.

Keywords: Optical Character Recognition; Handwriting; Image Processing; Pattern Recognition; off-line handwriting recognition;

1. INTRODUCTION

We ask that authors follow some simple guidelines. This Pattern recognition is the scientific discipline whose goal is the classification of objects into number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified.

The handwriting recognition refers to the identification of written characters. Handwriting recognition has been become a very important and useful research area in recent years for the ease of access of many applications. [1]

There are two types of handwriting recognition: off-line recognition and on-line recognition. Off-line handwriting recognition involves the automatic conversion of text in an image into letter codes which are usable within computer and text-processing applications. The data obtained by this form is regarded as a static representation of handwriting. Off-line handwriting recognition is comparatively difficult, as different people have different handwriting styles. On-line handwriting recognition involves the automatic conversion of text as it is written on a special digitizer or PDA, where a sensor picks up the pen-tip movements as well as pen-up/pen-down switching. This kind of data is known as digital ink and can be regarded as a digital representation of handwriting. The obtained signal is converted into letter codes which are usable within computer and text-processing applications.

2. ARABIC CHARACTERS

All Arabic characters are used in writing many languages not only in Arabic countries, but for Urdu and Farsi and other languages in countries where Islam is the principal religion (e.g., Iran, Pakistan, and Malaysia). [4] The special characteristics of Arabic written words and characters do not allow the direct application of algorithms for other languages. See figure 1.

خ	ح	ج	ث	ت	ب	أ
Xaa'	*H'aa'*	*Jeem*	*Thaa'*	*Taa'*	*Baa'*	*'Alif*
ص	ش	س	ز	ر	ذ	د
Saad	*Sheen*	*Seen*	*Zaay*	*Raa'*	*Thaal*	*Daal*
ق	ف	غ	ع	ظ	ط	ض
Qaaf	*Faa'*	*Ghayn*	*'Ayn*	*Thaa'*	*Taa'*	*Daad*
ي	و	هـ	ن	م	ل	ك
Yaa'	*Waw*	*Haa'*	*Noon*	*Meem*	*Laam*	*Kaaf*

Figure. 1 Arabic letters

Arabic's Letters characteristics are:

– Arabic is a cursive type language written from right to left.

– Arabic has 28 basic characters. Each character has 2-4 forms depending on its position within the word.

– Many letters of the Arabic alphabet have dots, above or below the character body, and some letters have a Hamza (zigzag shape) and dilation.

– Overlapping characters: some Arabic's characters become over each other horizontally when they connected with each other. [4]

Table1. Arabic characters and their shapes at different positions in the word

Letter	Stand-alone	Initial	Middle	Final	Other shapes
Alef	ا			ـا	
Ba'	ب	بـ	ـبـ	ـب	
Ta'	ت	تـ	ـتـ	ـت	ة ـة
Tha'	ث	ثـ	ـثـ	ـث	
Jeem	ج	جـ	ـجـ	ـج	
H'a'	ح	حـ	ـحـ	ـح	
Kha'	خ	خـ	ـخـ	ـخ	
Dal	د			ـد	
Thal	ذ			ـذ	
Ra'	ر			ـر	
Zai	ز			ـز	
Seen	س	سـ	ـسـ	ـس	
Sheen	ش	شـ	ـشـ	ـش	
Sad	ص	صـ	ـصـ	ـص	
Dhad	ض	ضـ	ـضـ	ـض	
Tta	ط	طـ	ـطـ	ـط	
Dha'	ظ	ظـ	ـظـ	ـظ	
Ain	ع	عـ	ـعـ	ـع	
Gahin	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'	ى	يـ	ـيـ	ـي	

3. OCR system

OCR systems consist of five major stages:

3.1 Image acquisition

Is the first step in the algorithm, the system takes the original image that needed to read, from the scanner or from computer storage, this process can be represented in follow chart as shown in Figure 2:

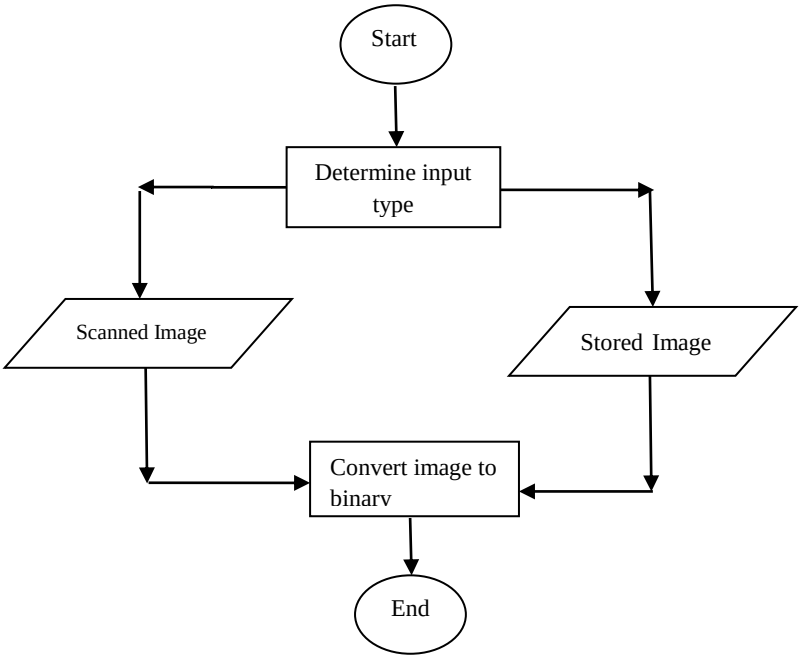

Figure. 2 Image acquisition algorithm

3.2 Pre-processing

The aim of preprocessing stage is the removal of all elements in the word image that are not useful for recognition process. It includes:

3.2.1 Binarization:

Binary images are the simplest type of images that take one of two values, typically black and white, or '0' and '1'. A binary image is referred to as a 1 bit/pixel image because it takes only 1 binary digit to represent each pixel. This type of image is most frequently used in computer vision application where only information required for the task is general shape, or outline, information. [3]

3.2.2 Smoothing:

Filling gaps and eliminating superfluous points of the contour image.

3.2.3 Cleaning:

Removing noise that could not be eliminated by smoothing.

3.2.4

Determine the size and refine the distortions of the image and their impurities which may be associated with it.

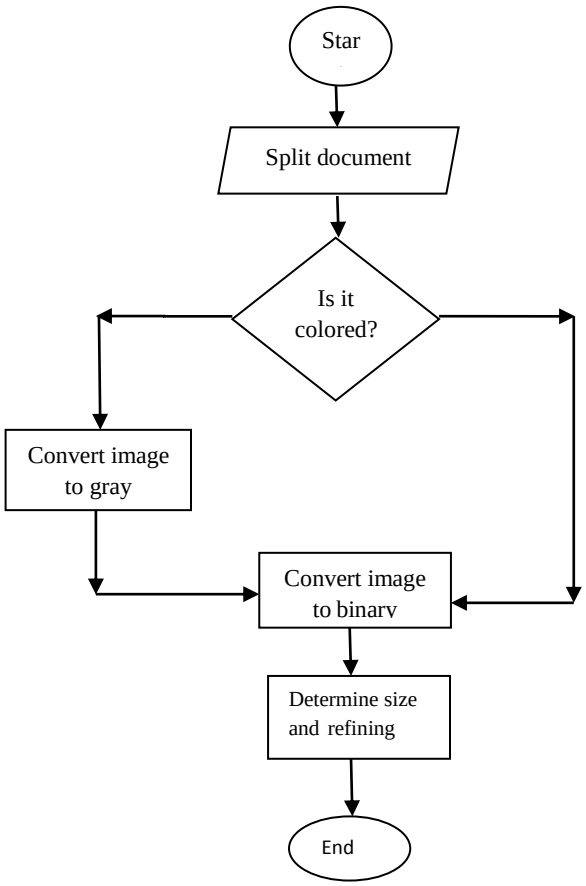

Figure. 3 Pre-process algorithm

3.3 Segmentation

3.3.1 First the document is divided into lines by using histogram, then calculating the number of dots in each horizontal pointed row.

3.3.2 Dividing the lines of the document into Characters, according to the shape of the Character, based on rules and information that owned by the system.

3.3.3 Extracting the features by collecting the dots in each row separately, and also to the columns, then studying and analyzing the characteristics such as Character height and width, in preparation to identify the Crafts.

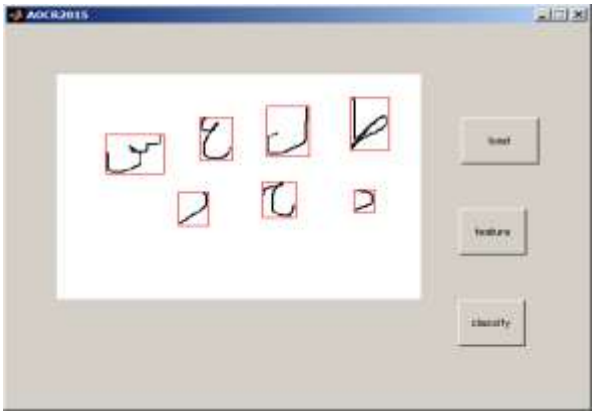

Figure. 4 image segmentation into lines and characters

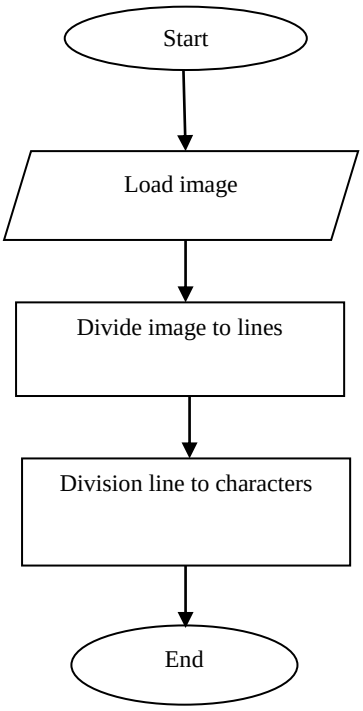

Figure. 5 Segmentation algorithm

3.4 Feature Extraction

This process extracts the features of the characters that are most relevant for classifying at recognition stage. This is an important stage as it can help avoid misclassification, thus increasing recognition rate.

In feature extraction stage each character is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

3.5 Classification

Classification is a variety of ways including linear correlation algorithm the linear correlation matrix phase is compared to the characters to be identified with matrices stored in the data base and are selected after comparison the largest correlation value is determined by the appropriate letter calculated the correlation between the value of the following formula:

$$x = \sum_m \sum_n \left(A_{mn} - \bar{A}\right)\left(B_{mn} - \bar{B}\right)$$

$$y = \sqrt{\left(\sum_m \sum_n \left(A_{mn} - \bar{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \bar{B}\right)^2\right)}$$

$$r = \frac{x}{y}$$

Where:

r = correlation value.

A = initial matrix (for character want to identify it).

B = template matrix (for stored character).

$\bar{A}$ = mean of the initial matrix (character want to identify it).

$\bar{B}$ = mean of the template matrix (for stored character).

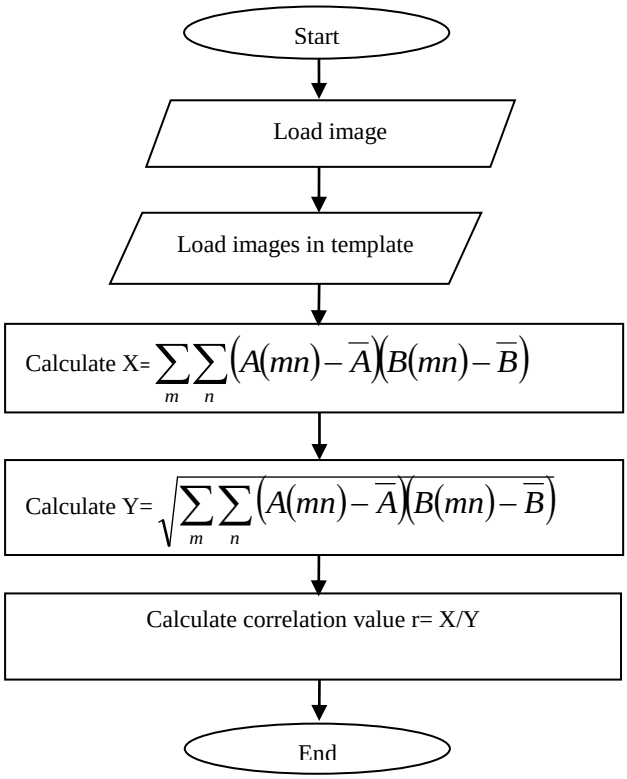

Figure. 6 linear correlation algorithm

At the end, the work of this algorithm can be summed up in the following chart:

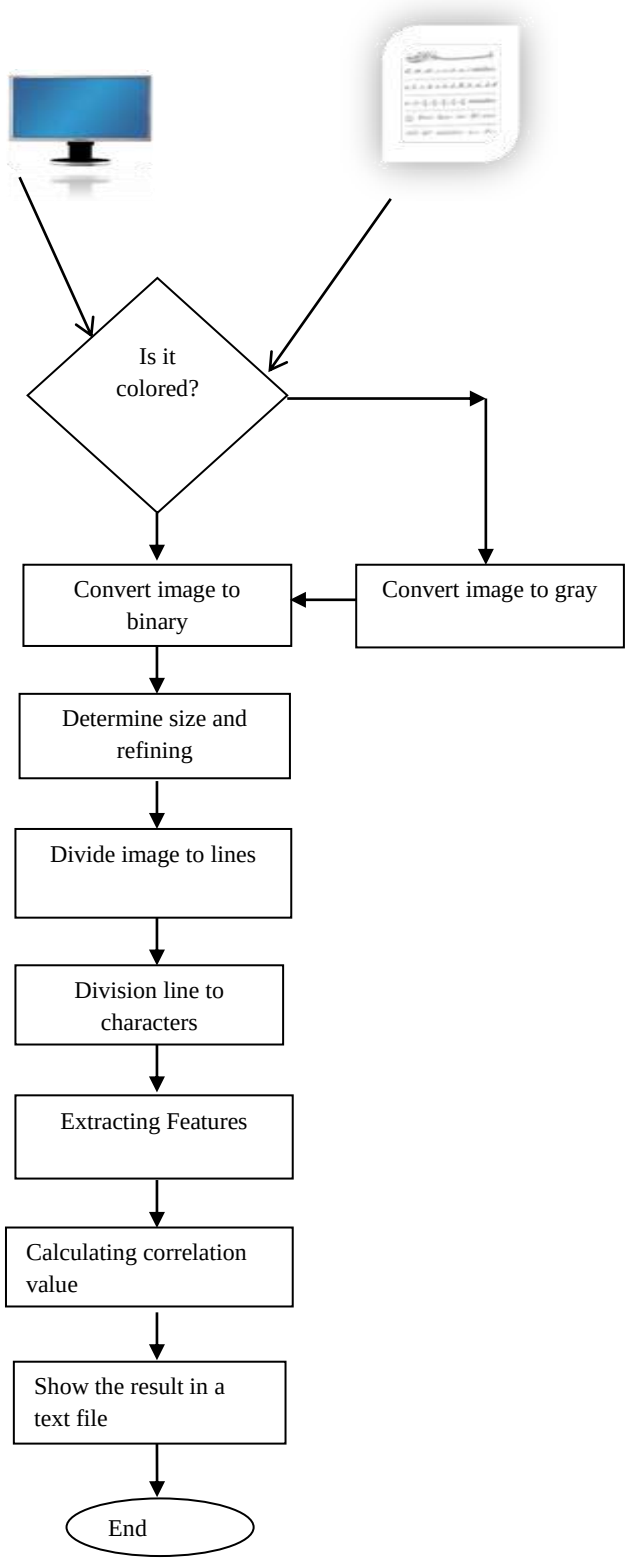

Figure. 7 : completely algorithm

3.6 Evaluating The Results

After the application of the proposed algorithm on the number of images we calculated the correlation factor and the ratio Signal to Noise Peak Signal-to-Noise Ratio (PSNR) between the input image and the resulting images, and is the PSNR account the following law:

PSNR = 10*log (255*255/MSE) / log (10)

The law used to calculate the mean square error (MSE)is:

MSE = sum (sum (error.* error)) / (M * N)

The results were as follows:

TABLE2. Recognition performances

no	Recognition performances of the PSNR	
	Character	PSNR
1	أ	32.0951
2	ب	29.8934
3	ت	33.0398
4	ث	36.6835
5	ج	36.6684
6	ح	35.8495
7	خ	35.8950
8	د	30.8357
9	ذ	33.9198
10	ر	32.8113
11	ز	35.2482
12	س	30.5137
13	ش	35.8866
14	ص	36.2440
15	ض	30.5039
16	ط	27.1389
17	ظ	28.4566
18	ع	35.1700
19	غ	30.5559
20	ف	34.8221
21	ق	31.1804
22	ك	37.2377

no	Recognition performances of the PSNR	
	Character	PSNR
23	ل	33.2289
24	م	28.6027
25	ن	36.1048
26	هـ	30.5709
27	و	36.7898
28	ي	34.7908

4. REFERENCES

[1] **Abdelhafeez hamid**, G. Zen Alabdeen, A. Mansour, "Arabic Numerals Recognition Using Linear Correlation Algorithm in Two Dimensions", International Journal of Computer Applications Technology and Research Volume 4– Issue 1, 01 - 06, 2014.

[2] K. Kaur1,N. K. Garg, "A Survey on Process of Isolated Character Recognition", International Journal of Innovative Research in Computer and Communication Engineering Vol. 2, Issue 5, May 2014.

[3] Haithem Al-Ani,Ban N,Dhannoon and Huda M. Abass, "Printed Arabic Character Recognition using Neural Network", Journal of Emerging Trends in Computing and Information Sciences, Vol. 5, No. 1 January 2014.

[4] Hasan Al-Rashaideh, "Preprocessing phase for Arabic Word Handwritten Recognition", Information Transmition In Computer Networks, 2006.

[5] Shatha M. Noor," Isolated Multifont Arabic Character Recognition Using Fourier Descriptors", Journal of Al-Nahrain University Vol.16 (1), March, 2013, pp.176-182.

[6] Ahmed M. Shaffie and Galal A. Elkobrosy, "A Fast Recognition System for Isolated Printed Characters Using Center of Gravity and Principal Axis", Applied Mathematics, 2013, 4, 1313-1319.

[7] M. A. Abdullah, L. Al-Harigy, and H. Al-Fraidi, "Off-Line Arabic Handwriting Character Recognition Using Word Segmentation", Journal of computing, Vol. 4, Issue 3, March 2012, ISSN 2151-9617.

[8] F. Parrweej, "An Empirical Evaluation of Off-line Arabic Handwriting And Printed Characters Recognition System", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 6, No 1, November 2012.

[9] B. Alijla and K. Kwaik, "Online Isolated Arabic Handwritten Character Recognition Using Neural Network", The International Arab Journal of Information Technology, Vol. 9, No. 4, July 2012.

[10] A. Pal and D. Singh, "Handwritten English Character Recognition Using Neural Network", International Journal of Computer Science & CommunicationVol. 1, No. 2, July-December 2010, pp. 141-144.

R. Kala, H. Vazirani, A. Shukla and R. Tiwari, "Offline Handwriting Recognition using Genetic algorithm", IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 2, No 1, March 2010 .

[11] M. Al-Alaoui, M. Amin, Z. Abou Chahine, and E. Yaacoub, "A New Approach for Arabic Offline Handwriting Recognition", IEEE Multidisciplinary Engineering Education Magazine, Vol. 4, No. 3, Semptember 2009.

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

[13] Mansoor Al-A'ali and Jamil Ahmad, "Optical Character Recognition System for Arabic Text Using CursiveMulti-Directional Approach", Journal of Computer Science 3 (7): 549-555, 2007 ISSN 1549-3636.