Al Azhar University Engineering Journal, April, 2001
Published by Faculty of Engineering , Al Azhar University, Cairo - Egypt

CONTOURS OF ARABIC CHARACTERS: A DIFFERENT POINT OF VIEW

Kamal Jambi and **Ibrahim Al-Bidewi**
King Abdul Aziz University, Jeddah, Saudi Arabia

ABSTRACT. This work draws attention to a different view over the contours of Arabic characters. It investigates different tests where the contours of Arabic characters are converted into graphs. Six different tests are implemented using Visual C++ under Windows environment. The tests are: ***X and Y*** graph, ***Radius*** graph, ***Theta*** graph, ***Chain Code-16*** graph, , ***Corners*** graph, and ***extremes*** graph of the letter. These tests were done on different characters. Some of the obtained graphs are shown. Also, features from these graphs can be easily used to classify Arabic characters.

1. INTRODUCTION

Although Arabic alphabet contains only 28 characters, yet the process of recognition deals with more than 60 characters. This is because Arabic characters take different shapes depending on their location within the word. Table-1 indicates the different shapes of the character when it is located at the start, middle, or end of the word. The character also has a different shape if it is isolated. Also, some characters are classified into some groups having the same main stroke with a minor change. This change is indicated by having different number of dots as well as their location with respect to the main stroke [1].

The set of features is defined to be the smallest set that can effectively discriminate between different classes. In a relatively recent survey [2], the authors classify features used in Arabic Character Recognition (ACR) into four groups. These are structural, statistical, global transformation, and template matching and correlation. Structural features deal with the topology and the physical structure of Arabic characters (i.e. loops, branch points, and end points) as implemented in [3, 4, 5]. Statistical features depend on the characteristic measurements such as crossings and moments as shown in [6, 7]. Global transformations alter the pixel representation of the character into a different type of representation such as Hough transform and Fourier descriptors. Template matching and correlation depend on comparing patterns such as comparing a set of connected pixels to a corresponding one.

Description of profiles is one of the fundamental problems in pattern recognition and has been generally accepted as one of the most important object attributes. This paper presents some experiments done by implementing six types of tests to deal with contours of Arabic characters. It should be mentioned that this work is a module of a system. The module should provide the system with a set of features to be used for classification by means of a neural network. In order to have a complete view regarding the following tests, the reader has to have an idea about the way each bit map of Arabic character is processed. The upper left corner is considered as the point of the origin and scanning of the contour is done in clockwise direction.

Contours features with respect to Arabic character recognition become popular. This is true since the contour is easy to find and quick to obtain comparing to the time-consuming process of thinning. Also, the contour absorbs characteristics of the Arabic characters. There are some applications of using contour in different manners. The most straight forward application is presented by using coordinates of the black pixels of the contour. Some of structural features of the contours such as convex or concave curves can be easily used as features of Arabic characters. Abo Samra and Jambi [8] used distances of feature points located in the contour. These distances are converted to some kind of structural feature. Another work of Jambi [9] made a transformation of x-y coordinates of pixels of contours to radial coordinates. This is followed by finding the distributions of radii within some predefined sectors. Another application is using the intersection of the contours with some predefined grid areas [10].

2. THE PROPOSED FEATURES

As mentioned earlier, features used for identifying Arabic Characters can be classified into four classes. Statistical features and structural ones are commonly used. Statistical features are those who are obtained by means of statistical measurements such as moments and Fourier descriptors. On the other hand, structural features depend on the physical structure of Arabic characters such as the existence of branch points, end points, corners, cups, loops, etc. Structural features are obtained from the skeleton of the words or characters through the process of thinning. However, thinning is a time consuming process and much noise sensitive. Therefore, researchers on this field moved towards analyzing contours of character. Moreover, the features to be looked after should be selected to be able to deal with these new trends. Over the last two decades, several different approaches to profile recognition problems have been investigated extensively. The attempts have been made to place profile characterization on a quantitative level by the use of various mathematical, statistical and linguistic descriptions.

Table 1. The different shapes of Arabic characters depending on their location within the word.

Name	Isolated	First	Middle	Last	Name	Isolated	First	Middle	Last
Alif	ا	ا	ا	ا	Dhad	ض	ضـ	ـضـ	ـض
Baa	ب	بـ	ـبـ	ـب	Tta	ط	طـ	ـطـ	ـط
Taa	ت	تـ	ـتـ	ـت	Zha	ظ	ظـ	ـظـ	ـظ
Thaa	ث	ثـ	ـثـ	ـث	Ain	ع	عـ	ـعـ	ـع
Geem	ج	جـ	ـجـ	ـج	Ghain	غ	غـ	ـغـ	ـغ
Hha	ح	حـ	ـحـ	ـح	Faa	ف	فـ	ـفـ	ـف
Kha	خ	خـ	ـخـ	ـخ	Gaf	ق	قـ	ـقـ	ـق
Dal	د	د	ـد	ـد	Kaf	ك	كـ	ـكـ	ـك
Thal	ذ	ذ	ـذ	ـذ	Lam	ل	لـ	ـلـ	ـل
Raa	ر	ر	ـر	ـر	Meem	م	مـ	ـمـ	ـم
Zain	ز	ز	ـز	ـز	Noon	ن	نـ	ـنـ	ـن
Seen	س	سـ	ـسـ	ـس	Haa	ه	هـ	ـهـ	ـه
Sheen	ش	شـ	ـشـ	ـش	Waw	و	و	ـو	ـو
Saad	ص	صـ	ـصـ	ـص	Yaa	ي	يـ	ـيـ	ـي

There are six types of tests to cover the following aspects: ***X and Y*** graph , ***Radius*** graph, ***Theta*** graph , ***Chain Code-16*** graph, ***Corners*** graph, and ***extremes*** graph of the letter. The following paragraph gives some explanation of these graphs.

2.1 X And Y Graph Of The Letter

This test is the most obvious one. The graph is obtained directly from the contour of the character as shown in figure (1). The graph contains two curves. One is corresponding to the X-coordinate and the other one is related to the Y-coordinate.

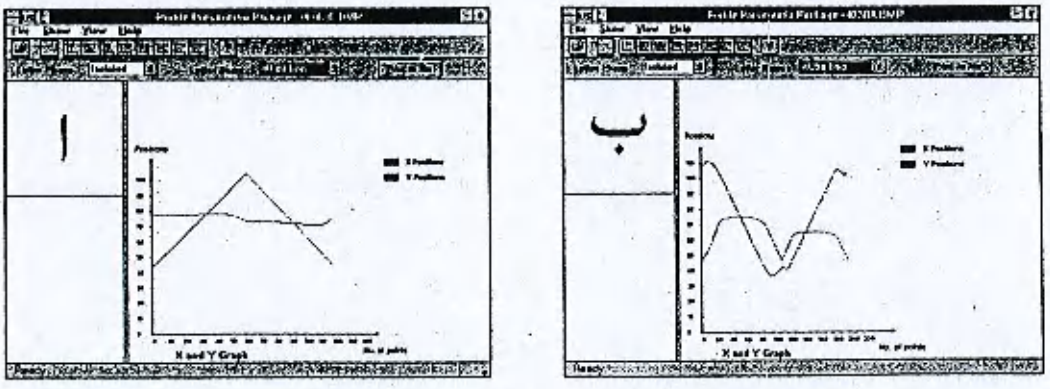

Figure 1: Representing some characters in ***X and Y*** graph.

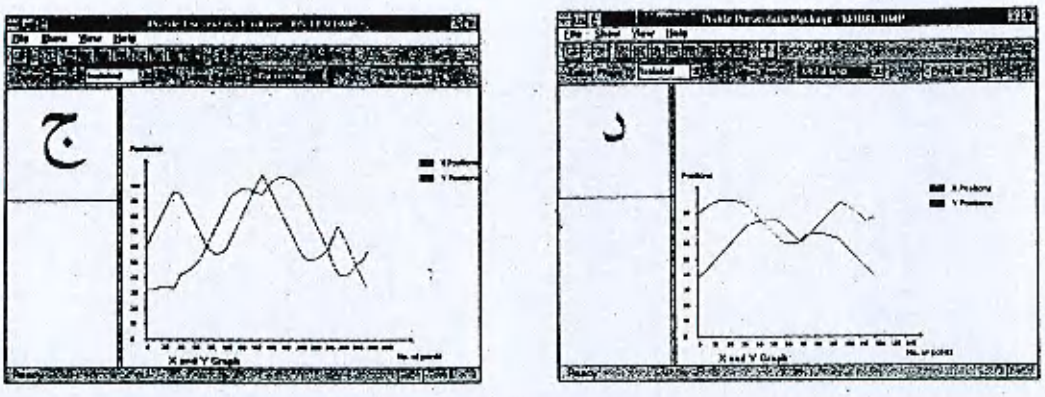

Figure 1(cont.) : Representing some characters in *X and Y* graph.

2.2 Radius Graph Of The Letter

The next obvious choice is to determine the center of the letter under investigation as seen in figure (2). This will be followed by plotting the graph of the distance of all pixels of the contour and that center. The coordinates of the center are calculated as:

x-value = 1/n $\sum$ i, for Image(i,j) $\in$ profile

y-value = 1/n $\sum$ j, for Image(i,j) $\in$ profile

2.3 Theta Graph Of The Letter

Theta is defined to be the external angle for each pixel in the contour of the letter. It can be determined by following the next steps:

Step 1 : Select a pixel (x,y) in the profile as the Center. It's coordinates are (center_x, center_y).

Step 2 : Determine another pixel located in a position after the ***center*** pixel. Let it be called ***after***. It has a fixed distance from the *center (e.g.* 10 pixels). It has the coordinate (after_x, after_y).

Step 3 : Determine another pixel located in a position before ***center*** pixel. Let it be called ***before***. It has the same fixed distance from the *center pixel*. It has the coordinate (befor_x, befor_y).

Therefore, connecting the above three pixels gives the shape of triangle as shown in figure (3). This is followed by some simple calculation to end up with the external angle.

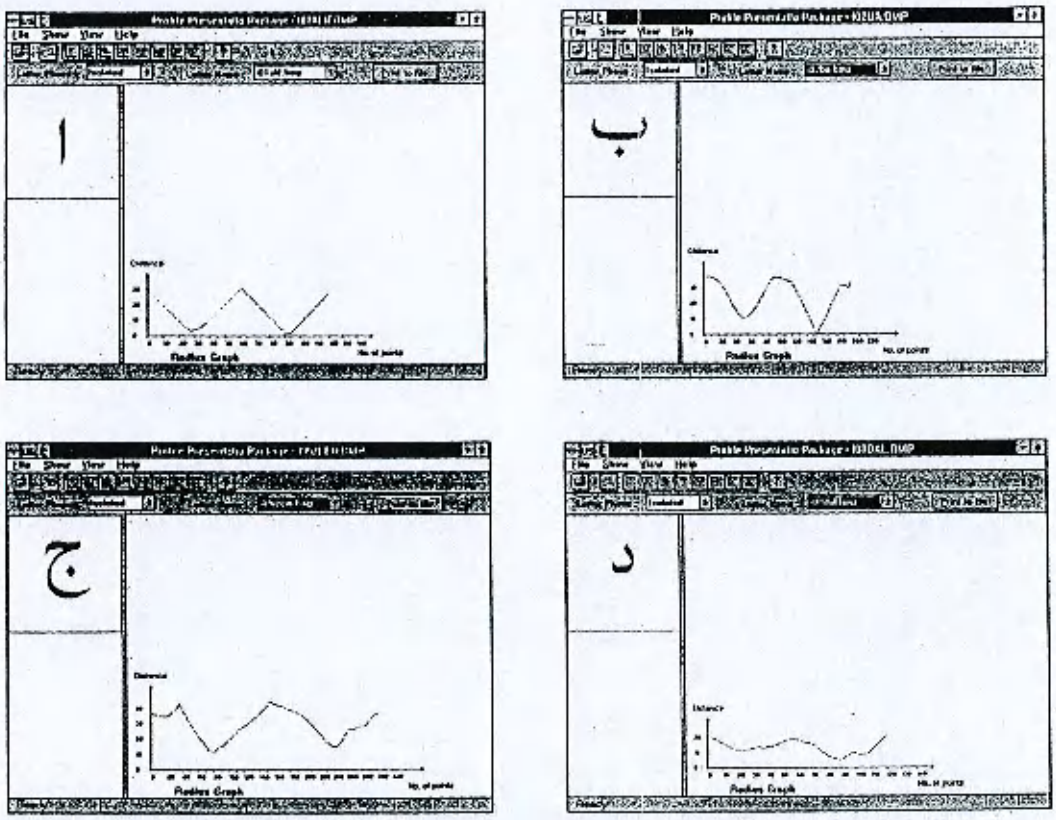

Figure 2 : Representing some characters in ***Radius*** graph.

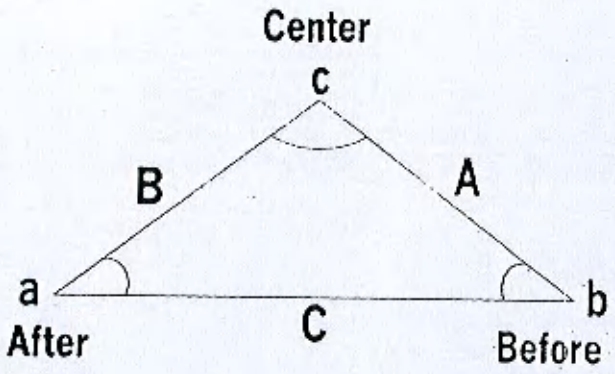

Figure 3 : Vertices of the triangle used to calculate *Theta*.

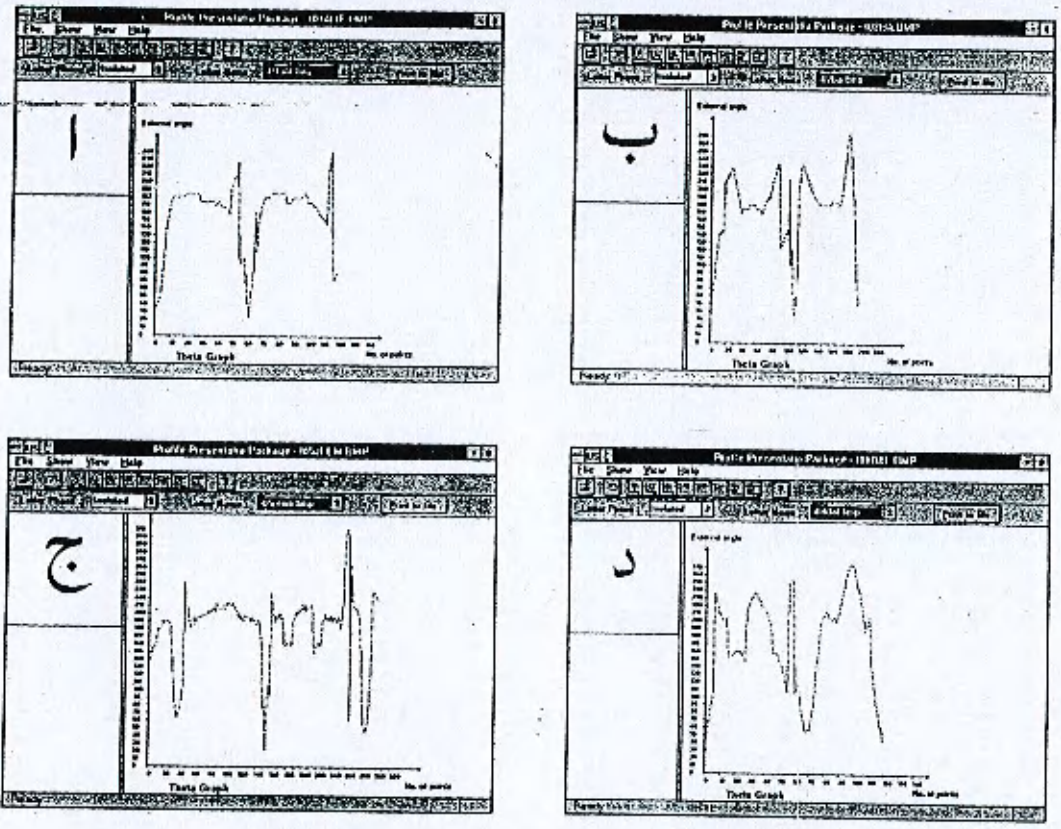

Figure 4 : Representing some characters in ***Theta*** graph.

2.4 Chain Code 16 Graph of the Letter

Chain codes are used to specify the direction of a contour segments. Directions are quantized into one of sixteen direction as shown in figure (5). Coding starts from the first segment and going clockwise around the contour. The direction to the next edge is determined by a code from the 16-chain codes as shown in figure (6).

2.5 Corners Graph of the Letter

The idea behind getting the graph of this feature is similar to the Radius figure (7). However, the main difference is represented in such a way that distances are calculated between each pixel of the contour and the other corners of the image of character. Corners are upper left, upper right, lower left , and lower right corners.

3	4	5	6	7
2	2	3	4	8
1	1		5	9
16	8	7	6	10
15	14	13	12	11

Figure 5 : The template used for chain codes 16.

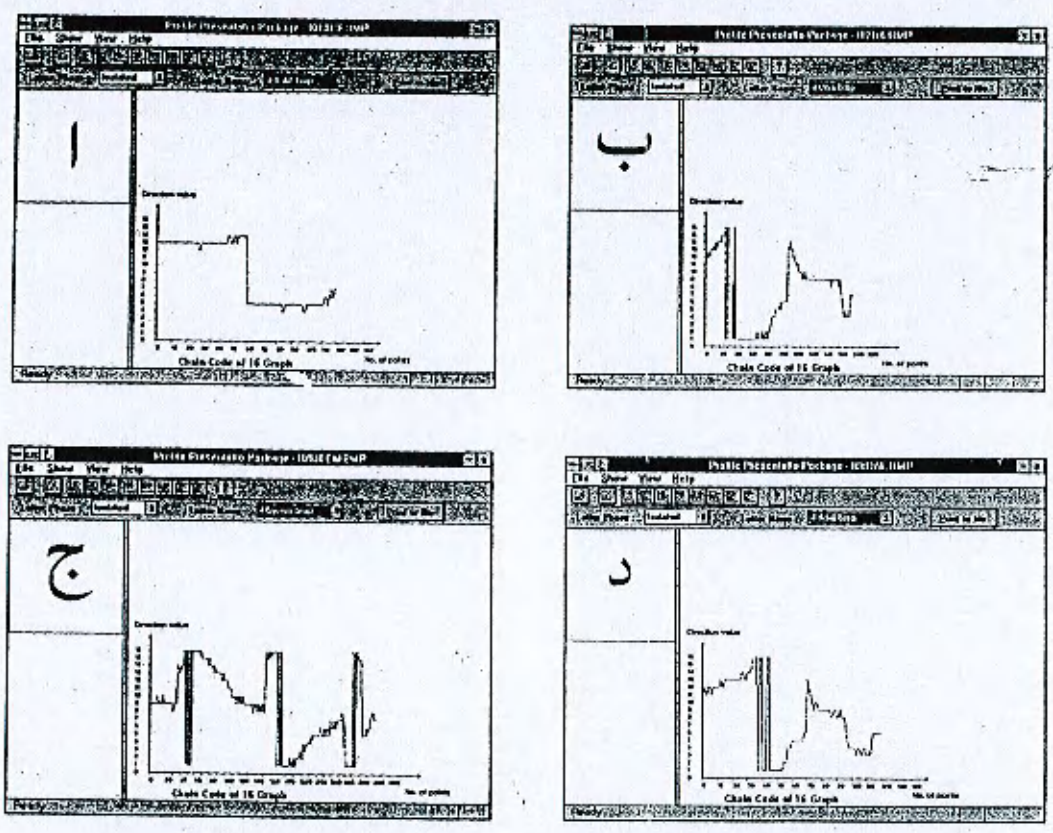

Figure 6 : Representing some characters in ***chain code 16*** graph.

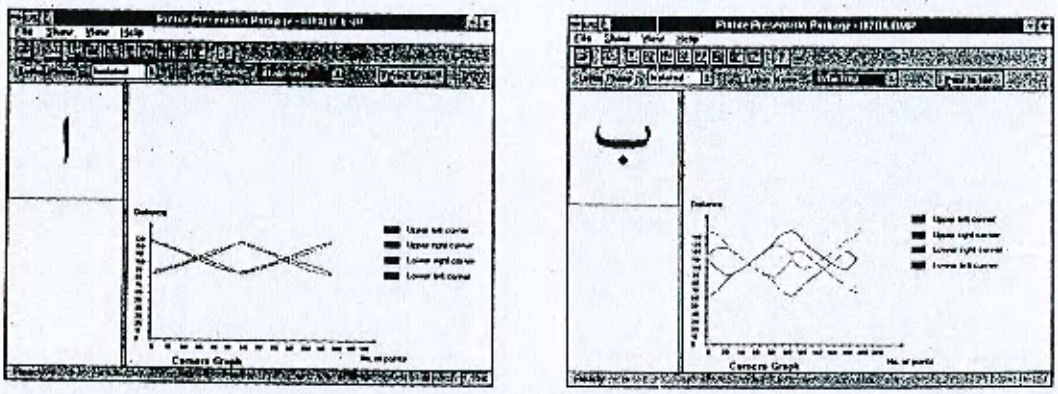

Figure 7 : Representing some characters in *Corners* graph.

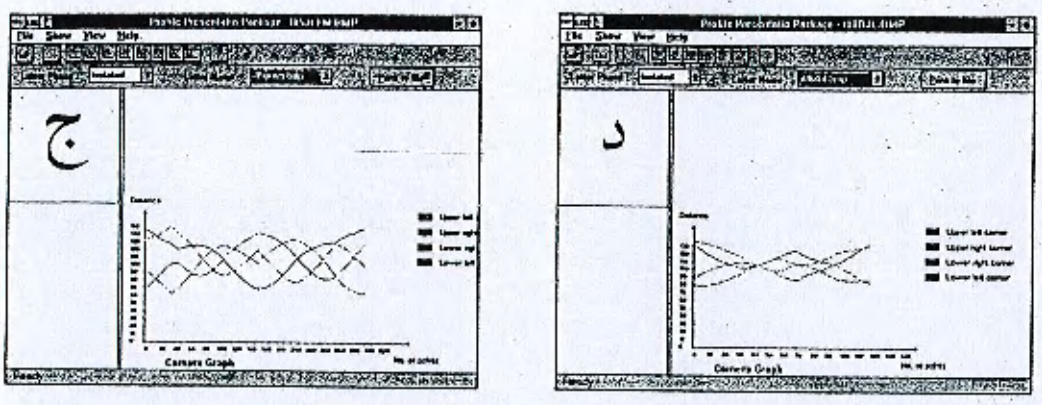

Figure 7 - (cont.) : Representing some characters in *Corners* graph.

2.6 Extremes Graph Of The Letter

The idea behind getting the graph is a repeated situation of the corners. The major difference concern the points where distances are calculated to the points of the profile. These points are the extremes in the direction of top, bottom, left and right.

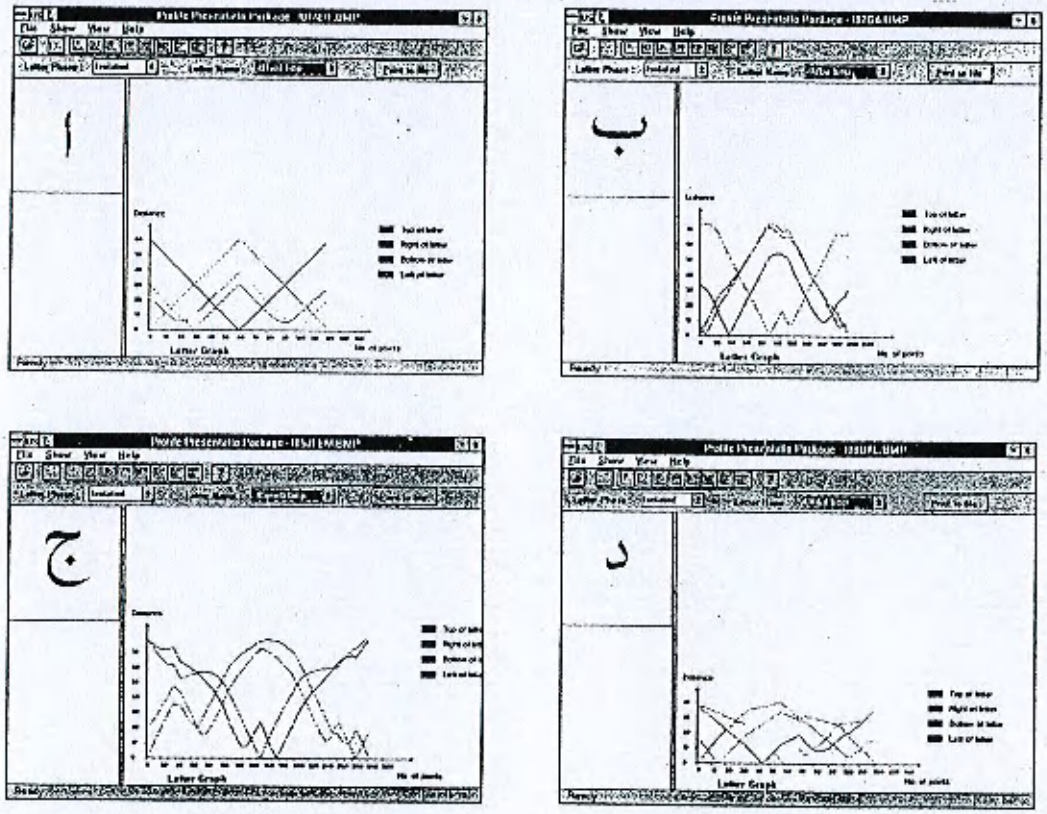

Figure 8 : Representing some characters in ***Extreme*** graph.

3. CONCLUSION

New representations of Arabic characters are contributed. They are promising in the sense that the obtained curves can be easily used to classify different types of Arabic characters. The classification process may be done by means of a neural network. Therefore, the corresponding values of these curves at some intervals hat should be fed into the neural network.

One major thing to keep in mind is that those features are contour based. This property makes these features preferable by current researchers in this field. This preference is due the simplicity of obtaining these contours comparing to the time consuming process of thinning. The ability to manipulate these curves gives the users the ability to reduce if not eliminate the effect of scaling, rotation and displacement of character image.

ACKNOWLEDGMENT

The authors of this work are very thankful to King Abdul Aziz City for Science and Technology. The appreciation is for the support of this work through the grant of project number (AT-15-20). The authors would like to thank Mr. Majid Bakkar for his comments and suggestions.

REFERNCES

[1] Jambi, K., "Recognition of Constrained Handwitten Arabic Words Without Segmentation", Proceedings of the 6th International Conference on Computer Theory and Applications, Alexandria, 3-5 Sept., 385-395, 1996.

[2] Al-Badr, B., and Mahmoud, S., "Survey and Bibliography of Arabic Optical Text Recognition", Signal Processing, **Vol. 41**, 49-77, 1995.

[3] Jambi, K., "An Approach for Segmenting Handwritten Arabic Words", *Proceeding of The Second Conference on Arabic Language*, Casablanca, Morroco, 233-243, 1993.

[4] Jambi, K., "An Experimental Approach for Recognizing Handwritten Arabic Words", Jornal of King AbdulAziz University Science, **Vol. 7**, 119-130, 1995.

[5] Almuallim, H., and Yamguchi, S., " A Method of Recognition of Arabic Cursive Handwriting", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **Vol. 9, No. 5**, 715-722 1987.

[6] El-Dabi, S., Ramsis, R., and Kamil, A., "Arabic Character Recognition System: A Statistical Approach for Recognizing Cursive Typewritten Text", *Pattern Recognition*, **Vol. 23, No. 5**, 485-495 1990.

[7] Al-Badr, B. and Haralick, R., "Recognition without segmentation: Using Mathematical Morphology to Recognize Printed Arabic", *Proceedings of The 13th National Computer Conference*, 813-829, 1992.

[8] Abo Samra, G., and Kamal Jambi"Using Contour-Based Features for Segmenting and Recognizing Handwritten Arabic Words", Proceeding of the IASTED International Conference Signal and Image Processing, October 27-31, 1998, Las Vegas, Nevada - USA.

[9] Jambi, Kamal, "An Approach for Arabic Signature Recognition", Sent for Publication.

[10] Bajaj, R. and S. Chaudhury, "Signature Verification Using Multiple Neural Classifiers", Pattern Recognition, **Vol. 30**, No. 1, pp. 1-7, 1997.