Arabic Character Recognition Using Fractal Dimension

Khalil I. Alsaif　　　　　　　　**Karam Hatim Thanoon**

khalil_alsaif@uomosul.edu.iq　　　*karamhatim@uomosul.edu.iq*

College of Computer Sciences and Mathematics

University of Mosul, Iraq

ABSTRACT

In this work the concepts of the pattern recognition was used to recognize printed Arabic characters, and the Fractal geometric dimension method was used.

The input for the system is image, with bitmap format , then the image of character is recognized, and after that it is feeding to the OCR system. A feature space containing the values of the fractal dimension for the letters of Arabic was constructed. These features were used in the recognition phase. In this phase a comparison was made between the values in the feature space and the values of the letter inputs to be recognized, the comparison was done by the minimum Euclidian distance. Results of this work are 75% succeeded. And Matlab 6.5 is used to write the functions and subroutines for this work.

Keywords: Fractal Dimension, optical character recognition, feature extraction**.**

تمييز الحرف العربي بأستخدام البعد الكسري

أ.م.د خليل ابراهيم السيف　　　م.م كرم حاتم ذنون

كلية علوم الحاسوب والرياضيات / جامعة الموصل

تاريخ الاستلام: 2008/10/13　　　　　　تاريخ القبول: 2008/12/04

الملخص

في هذا البحث استخدام مفهوم تمييز الأنماط لغرض تمييز الحروف العربية المطبوعة. تم استخدام طريقة أبعاد الهندسة الكسرية في هذا العمل.يتم ادخال الصورة باستخدام الماسح الضوئي وتكون من نوع bmp ، وبعد ذلك تغذى هذه الصورة الى نظام تمييز الحرف الضوئي (OCR). استخدم في البحث ابعاد الهندسة الكسرية لاستخلاص الخواص الخاصة بالحرف العربي، وهذه الخواص هي التي تستخدم في طور التمييز ، حيث يتم عمل مقارنة بين قيم الخواص المستخلصة وقيم الحرف المدخل لغرض التمييز وتمت هذه المقارنة باستخدام المسافة الاقليدية الصغرى. تم الحصول على 75% من النتائج الصحيحة واستخدمت لغة Matlab 6.5 لبناء الدوال والبرامج الفرعية لهذا البحث.

الكلمات المفتاحية: البعد الكسري، تمييز الحروف، استخلاص الخواص.

1. Introduction :

The word fractal dimension was used at the first time in the geometry area by the scientist Mandelbrote, which is made a big changes in the area of dealing with the natural images that is containing natural parts like trees, mountains, clouds, cost borers, … etc [10]. The calculation of fractal dimension is done by depending on self similarity, which is depending on dividing a shape into many parts, each part is similar to the base shape, and the process of dividing to smaller parts gives the same results. This means that the image is the same if the vision measure is changed [10].

For understanding the fractal dimension in a better way, specially in the area of images, it is important to mention topological dimension [4]. The topological dimension for a strait line or curve is one, for two dimensional shapes or surfaces is two, like square and rectangle, and for three dimensional shapes like cylinder or cube is three.

But the topological dimension for strait lines and curves in the natural shapes like borders, mountains, clouds, trees, or sky is very difficult to determine its dimension, because of the difficulty of using normal geometry to determine the dimensions for this type of shapes, therefore the fractal dimension are used for these types of shape [1].

2. Image Segmentation :

Image segmentation is counted as one of the most important steps that are used in image analysis for the purpose of shape recognition, and this is done by recognition of symmetric area in the image according to some properties of symmetry. Image segmentation is counted as the critical step in almost all computer vision problems, because there exists many factors that effect the quality of the segmentation, that is selecting the proper algorithms for feature extraction in addition of classification algorithms [2].

Fractal is one of the good ways for constructing and analysis of natural texture. Many researchers used fractals directly as measure for texture. The estimation of fractals is done either through frequency domain by estimating degree of curve deviation of the logarithm of power stream, or through spatial domain by using different methods like, box counting algorithm or two dimension variation algorithm for the calculation of fractal dimension.

The Koch Curve [see fig.1] shows the way of drawing a picture by using fractal dimension, at the first the straight line is segmented into three equal parts, the middle segment is exchanged by two segments that their length is the same as cutting size, so they get four segments with same size, each one is 1/3 total length of the segmented line. Second each segment is exchanged by three segment and so on, by a few steps they get a complicated shape by using simple formats [10].

The method that is used for calculating fractal dimension is Hausdroff-Bescovitch dimension, which is the simplest method for fractal dimension calculation [9]. In this method the image is divided into parts, the size of each of them is (r) and N(r) is the number of parts to cover the image [4].

the simplest way to calculate the volume of any element is by cutting the element into cubs [see fig.2] of volume (r), then calculates the number of cubes to cover the element N(r).
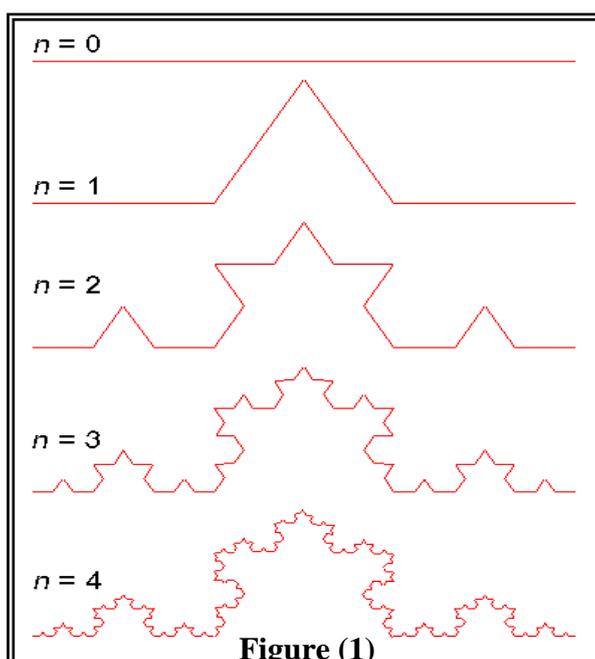
Figure (1)
Koch curve
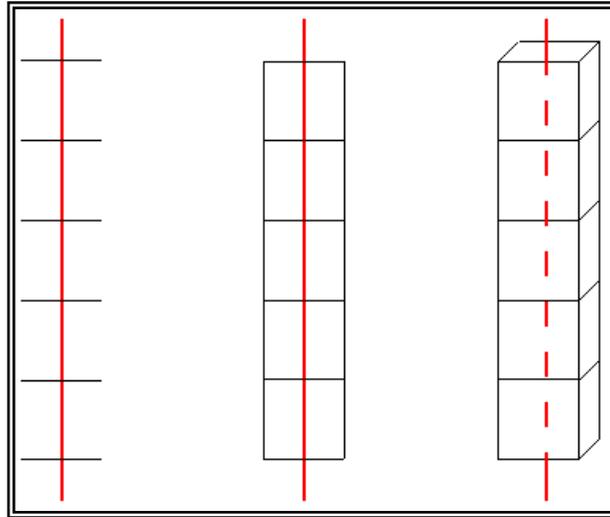
Figure (2)
Calculating strait line size

The length of a strait line is calculated by the following equation:

$N(r) = L_0 / r$ ...(1)

Where N (r): is the number of the segment got .

L$_0$: is the line total length , r :is the size of the segment.

Thus the length of the line equal to L$_o$ according to the following formula:

$$L = \lim_{r \to 0} N(r).r = \lim_{r \to 0} L_0 r^O = L_0 \qquad \ldots(2)$$

L will be approximately equal to the length of the line and independent on r [6]. Then for matching the area, the number of squares that is recoded to cover the points on the line is N(r) and the area of each square r^2, the area A is calculated as follows :

$$A = \lim_{r \to 0} N(r).r^2 = \lim_{r \to 0} L_0.r^1 = 0 \qquad \ldots(3)$$

And the volume V as follows :

$$V = \lim_{r \to 0} N(r).r^3 = \lim_{r \to 0} L_0.r^2 = 0 \qquad \ldots(4)$$

Both A and V go towards zero when (r) goes to zero, therefore the important measure is the length of the line.

To calculate the surface size [see fig.3], the normal measure is the area A where :

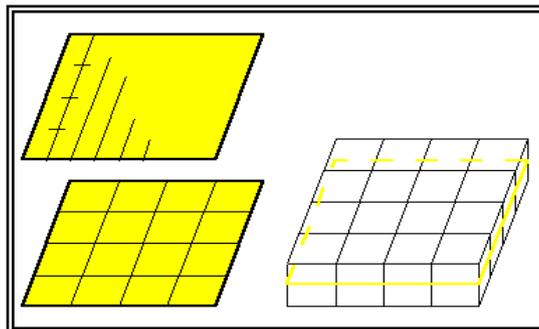$$A = \lim_{r \to 0} N(r).r^3 = \lim_{r \to 0} A_0.r^0 = 0 \qquad \ldots(5)$$

The number of the squares required to cover the surface is N(r) [4].

And the volume is calculated by :

$$V = \lim_{r \to 0} N(r).r^3 = \lim_{r \to 0} A_0.r^1 = 0 \qquad \ldots(6)$$

By this, it is possible to connect the length with surface where :

$$V = \lim_{r \to 0} N(r).r^3 = \lim_{r \to 0} A_0.r^1 = 0 \qquad \ldots(7)$$

Figure (3)
Surface size

The length deviates when (r) is approaching zero, by this way, the best measure for the surface is the area, and for the fractal dimension the volume is generalized by depending on the following function :

h(r) =y.r^d, and used to cover the surface (S) to form the measure M(d) = □h(r), where r is geometric factor [4], generally the value of M(d)is 0 or infinity when (r) gets close to zero depending on the value of (d), Hausdroff-Becovith dimension for (S) is the critical value for (d), when M(d) is changed from zero to infinity [3].
Where :

$$M_d = \sum \gamma.r^d = \gamma.N(r).r^d \xrightarrow{\quad r \to 0 \quad} \begin{cases} 0, d > D_H \\ \infty, d > D_H \end{cases} \qquad ...(8)$$

For Koch curve [see fig.4], the line is divided into four equal parts each of which of length 1/3 (first the straight line is segmented into three equal parts, the middle segment is exchanged by two segments that their length is the same as cutting size, therefore it will be four equal part) where N=4, r=1/3 and for the calculation of Hausdroff-Becovith dimension the following equations are used [5] :

$$N = r^d \qquad \qquad \qquad ...(9)$$

$$d = -\frac{\log N}{\log r} \qquad \qquad \qquad ...(10)$$

The value of the fractal dimension for Koch curve is as follows :

$$d = \frac{\log 4}{-\log 3} \cong -1.26 \qquad \qquad ...(11)$$
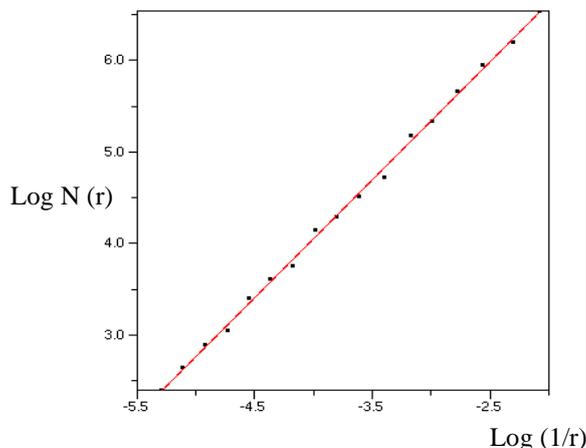
Figure (4)
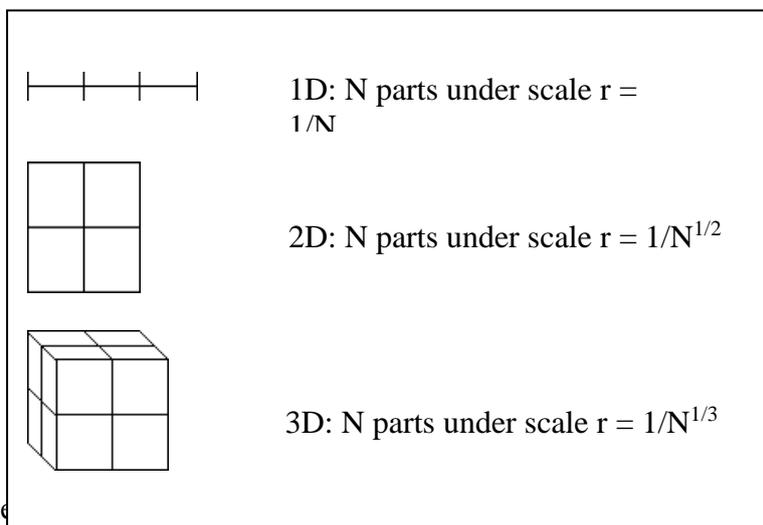**Slope of Koch curve**

3. Self Similarity :

After calculating the fractal dimension using Hausdroff-Becovith dimension, now it is possible to define the value fractals as the distance between Hausdroff-Becovith dimension and topological dimension, this definition is only for the field of dealing the natural part of images, the property of self similarity is the most important property of fractals, that can be visualized in many examples like cost borders, Koch curve, mountains, clouds, …. etc.

For studying property of self similarity in the shapes of one dimension like straight line, then dividing the line into (N) of equal parts, size of each of them is (r) [see fig.5].

Where :

$$r = \frac{1}{N} \qquad \qquad …(12)$$

Where (N) is the total number of segments.

1D: N parts under scale r = 1/N

2D: N parts under scale r = 1/N$^{1/2}$

3D: N parts under scale r = 1/N$^{1/3}$

For the square shape dividing able it into (N) of equal parts, Where :

Figure (5) Using the same measure to segment (1 dim ,2 dim ,3 dimension)

$$r = \frac{1}{\sqrt{N}} \qquad \qquad …(13)$$

For the three dimensional shapes like cube the shape is divided into equal parts where :

$$r = \frac{1}{\sqrt[3]{N}} \qquad \qquad … (14)$$

The property of self similarity for elements, where the elements are divided to (N) similar elements, all pieces have the same size according to the factor (r) for the element, the fractal dimension is as follows :

$$r = \frac{1}{\sqrt[d]{N}} \qquad \qquad … (15)$$

$$N = \frac{1}{r^d} \qquad \qquad … (16)$$

When similar elements (N) found using (r) for the element completely the fractal dimension is as follows :

$$d = -\frac{\log N}{\log r} \qquad \qquad …(17)$$

4. The Fractal Geometry and Self Similarity :

The fractal geometry depend on the property of self similarity, that means, if a shape is divided into number of parts, each part is similar to base shape, a predetermined scale is used to segment the shape into parts depending on the scale and the number of parts which is got the calculation of fractal dimension for the shape [10]. There are many methods for calculate the fractal dimension [8] :

1. Box counting method.
2. Two dimension variation method.

4.1 Box counting method :

Box counting method is the easiest and simplest method for calculating the fractal dimension for binary images [4], this algorithm based on dividing the image into number of squares of equal size (r) (which represent the number of pixels in each square), and this after changing the dimension of the image to be dividable, then calculate the number of squares that is needed to cover the shape, and calculating the logarithm of the number for the squares per the logarithm of (r) as follows :

$$d = -\frac{\log N}{\log r} \qquad \ldots (18)$$

This process is repeated with changing the value of (r), then calculate the dimension again, this is stopped when there is no changes in the value [10].

4.2 Two dimension variation :

This method is the best method for calculating the fractal dimension for the gray scale images according to the factor (r) [4], calculating the fractal dimension using this method, the center of the image is determined using a predetermined scale and defining value (r), then calculating the scale of each square and calculating the logarithm of it, these steps are repeated and changing the scale of (r), where r can be [3,5,7,9,…] the logarithm N(r) and logarithm of (1/r), then (D) is calculated according to the following equation :

$$D = 3 - \frac{F}{2} \qquad \ldots (19)$$

5. Implementation details

When we study the result of our study about some character we find that we must know this points which help us to recognize the character in a good manner .these points stored in a data base which useful in our work ,These points must be considered to help us to recognize the character, these points are:

1. the character .
2. Style of character .
3. Font of character.
4. Size of character .
5. Size of the picture of the character before and after treatment .
6. Number of pixels before and after treatment .
7. The picture of the character after truncation .
8. The picture of the character after thinning .
9. The value of fractal dimension .
10. Number of iteration .

This work have three main steps :
1- Truncation

The picture of the character after we cut all the spaces in it give a good results for recognition ,So this table explain that :

Table (1) the effect of truncation about fractal dimension for the letter(ب)

Fractal (after truncation)	Fractal (before truncation)	Size	font	character
1.08264	0.55185	10	Simplified Arabic	ب
1.18253	0.754255	12	Simplified Arabic	ب
1.20767	0.808289	14	Simplified Arabic	ب
1.2221	0.87449	18	Simplified Arabic	ب
1.24842	0.921605	24	Simplified Arabic	ب

Table (2) the effect of truncation about fractal dimension for the letter(ن)

Fractal (after truncation)	Fractal (before truncation)	Size	font	character
1.04421	0.49333	10	Simplified Arabic	ن
1.11695	1.00377	12	Simplified Arabic	ن
1.12986	0.373082	14	Simplified Arabic	ن
1.16215	0.863197	18	Simplified Arabic	ن
1.2377	0.99117	24	Simplified Arabic	ن

2- Thinning

The picture of the character after we reduce the pixels in it give a good results for recognition ,So this table explain that :

Table (3) the effect of thinning about fractal dimension for the letter(ب)

Fractal After thinning	Fractal After truncation	size	font	character
1.32205	1.08264	10	Simplified Arabic	ب
1.37637	1.18253	12	Simplified Arabic	ب
1.34473	1.20767	14	Simplified Arabic	ب
1.3666	1.2221	18	Simplified Arabic	ب
1.39407	1.24842	24	Simplified Arabic	ب

**Table (4) the effect of thinning about fractal dimension for the letter(ن)**

Fractal After thinning	Fractal After truncation	size	font	character
1.29777	1.04421	10	Simplified Arabic	ن
1.34888	1.11695	12	Simplified Arabic	ن
1.34944	1.12986	14	Simplified Arabic	ن
1.32736	1.16215	18	Simplified Arabic	ن
1.38983	1.2377	24	Simplified Arabic	ن

3- compute fractal dimension

after we do above operations we find the value of fractal dimension using box counting method and then compare this value with values stored in the database ,then know the character. The table (5) and the figure (6) show that the value of fractal dimension was stable when the size of character was variant ,which give a study about the relation between the fractal dimension and the size of character and show the number of iterations needed to compute fractal dimension.

Table (5) the relation between the fractal dimension and the size of character

Number of iterations	Fractal dimension	size	character
4	1.12271	10	ت
4	1.19613	14	ت
5	1.1822	18	ت
5	1.18392	22	ت
5	1.18024	28	ت

176
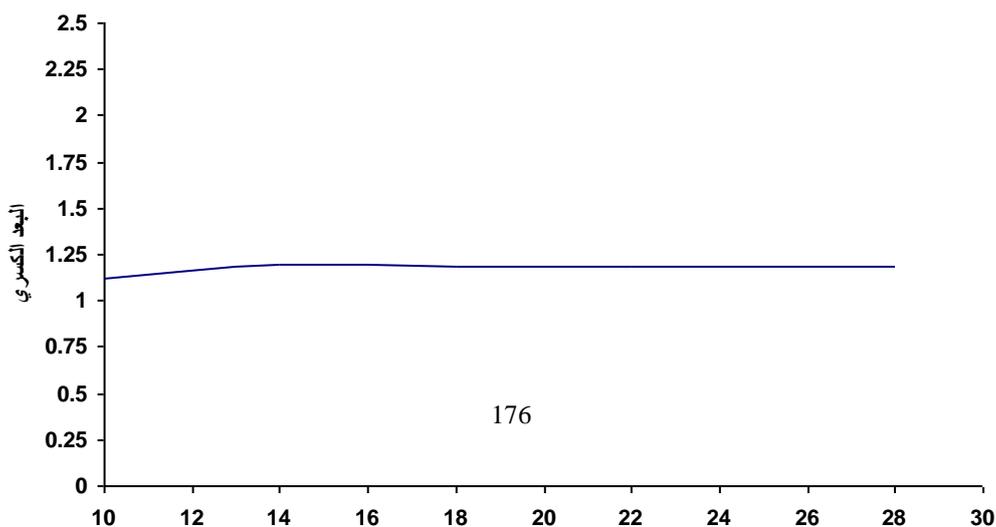
Figure (6)
The stability of fractal dimension when the size was variant

Conclusion:

In this research an algorithm based on the value of fractal dimension was applied as a feature of printed Arabic character ,we find that we can use fractal dimension to recognize printed Arabic character. There are many operation we do to got our results such as truncation and thinning, the truncation operation reduce the time of the recognition operation and thinning operation give the ability for high recognition and reduce the number of iteration until finished compute the fractal dimension. Practical application of our research provide high stability for the value of fractal dimension when the size of character was variant. for the future work we may connect the values of fractal dimension with a neural network which itself do the recognition operation.

REFERENCES

[1]   Gulick, D. (1992). "Encounters with Chaos", McGrew – Hill Inc.

[2]   Amin, A. (2000). "Recognition of Printed Arabic Test Based on Global Features and Decision tree Learning Technique", Pattern Recognition, Vol. 33, pp. 616-623.

[3]   Kim, Steenstrup and Mads Nielsen (2000). "The Housdroff Dimension and Scale-Space Normalization of Natural Images", Journal of Visual communication and image representation, Vol. 11, 266-277.

[4]   Kraft, (1995). "Estimating the Fractal Dimension from Digitized Images", Munich University of Technology, Statistics and Data Processing Institute, D-85350 Freising, Germany.

[5]   Mahadevaswamy, H.R. (2000). "New Approaches to Image Compression", Ph.D. Thesis, Department of Electronics Engineering, Regional Engineering College, The University of Calicut, Calicut, Kerala.

[6]   Mandlbroat, B.B. (1982). "The Fractal Geometry of Nature", Freeman, W.H. and Company, New York, U.S.A.

[7]   Narushige Shiode (1998). What is fractal. Center of advance spatial analysis, Theory and Application of GIS, Vol. 6, No. 1, pp. 81-86.

[8]   Neary, D. (2002). "Fractal Method in Image Analysis and Coding", Ms.C. for Dublincity University, Dublin, Iralnd.

[9]   Robert, P., Munafo (1999). "Hausdroff Dimension", Via internet.
       www.mandelbrotglossaryandencyclopedia

[10]   Voss, R.F. (1988). "Fractals in Nature: from characterization to simulation", H.O. Peitgen and D. Saupe, editors, The science of fractal images, Springer Verlag, New York, 1st edition, pp. 21-70.