A Study on the Fractal Attractor Creation and Analysis of the Printed Korean Characters

Young-Woo Shon, *Member, KIMICS*

Abstract—Chaos theory is a study researching the irregular, unpredictable behavior of deterministic and non-linear dynamical system. The interpretation using Chaos makes us evaluate characteristic existing in status space of system by time series, so that the extraction of Chaos characteristic understanding and those characteristics enables us to do high precision interpretation. Therefore, This paper propose the new method which is adopted in extracting character features and recognizing characters using the Chaos Theory. Firstly, it gets features of mesh feature, projection feature and cross distance feature from input character images. And their feature is converted into time series data. Then using the modified Henon system suggested in this paper, it gets last features of character image after calculating Box-counting dimension, Natural Measure, information bit and information dimension which are meant fractal dimension. Finally, character recognition is performed by statistically finding out the each information bit showing the minimum difference against the normalized pattern database. An experimental result shows 99% character classification rates for 2,350 Korean characters (Hangul) using proposed method in this paper.

Index Terms— Printed Korean characters recognition, Pattern Recognition, Image processing, Chaos Theory, Strange Attractor, Fractal Dimension, features extracting

I. INTRODUCTION

IN the past two decades, the world of science has come to a fundamentally new understanding of the dynamics of phenomena that vary over time. Grounded in mathematical discovery, yet given empirical substance by evidence from a variety of disciplines, this new perspective has led to nothing less than a reexamination of the concept of the predictability of dynamic behavior. Our implicit confidence in the orderliness of dynamical system, specifically of nonlinear dynamical systems, has not, it turns out, been entirely justified. Such systems are capable of believed. Fittingly, the new ideas are said to concern chaotic dynamics, or, simply, chaos[1-5]. Therefore, this paper proposes the new method which is adopted in extracting character using fractal dimension of Chaos theory which highly recognizes a minute difference with strange attractor created from Henon system. This paper implements a high precision character recognition system.

In recognizing Hangul character (Korean characters), the same kind of difficulties as found in English word recognition should be considered because Hangul character has a similar hierarchy graphemes are combined to form a character.

In order to solve the variability and the ambiguity problems, statistical and structural methods have been mainly used in previous studies[7-11]. The statistical method is basically data-driven in that the probabilities of given input patterns are estimated for recognition. The structural method, on the other hand, is a knowledge-driven method which examines the existence of characteristic components and their relationships. Although statistical methods have a well-defined mathematical foundation, assumptions sacrificing correctness are required to reduce the computational complexity. Statistical methods also tend to ignore small intentional differences, such as found in the '**q**' and '**g**' problems [7]. On the other hand, structural methods are capable of characterizing hierarchically embedded structural patterns. However, it is hard to build descriptions automatically.

This paper is organized as follows. In Section 2, a Structure of Hangul character is described. In Section 3, a Proposed Character Recognition system is described and the creation of character attractor is described. Implementation issues and experimental result are given in Section 4.

II. STRUCTURE OF HANGUL CHARACTER

Hangul character has a two-layered architecture:a grapheme layer and a character layer. The grapheme layer defines the kinds and shapes of graphemes, and the character layer defines character composition rule. Hangul character consists of 24 fundamental graphemes : 10 vowels and 14 consonants. Additional consonants and vowels are created by combining the fundamental graphemes (See Table 1).

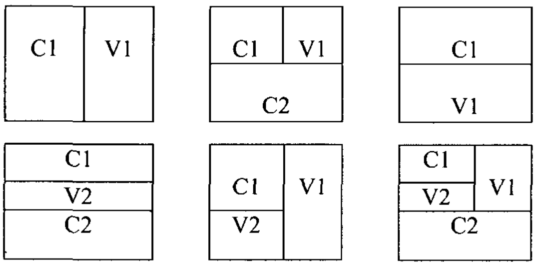
TABLE I The grapheme categories of Hangul character

Type	**Basic elements of Hangul**
Consonants (14)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ ㅈ ㅊ ㅋ ㅌ ㅍ ㅎ
Vowels (10)	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ

Manuscript received January 17, 2003.

Young-Woo Shon is with the Dept. of Computer Science, Kimpo College, 14-1, Ponae-ri, Wolgot-myun, Kimpo, Kyonggi-do, 415-873, Korea.
(Tel:+82-31-999-4209, FAX:+82-31-999-4775, e-mail: ywson@kimpo.ac.kr).

The graphemes are combined together to form a character by simple and logical rules. There are two types of vowels in hangul character : vertically elongated and horizontally elongated. The vertically elongated vowels have consonants on the left and the horizontally elongated vowels have them on the top.



C1: first consonant, C2: final consonant
V1: vertical vowel, V2: horizontal vowel

Fig.1 The structure of a Hangul character. (Six types of grapheme combination)

If a character has a consonants following a vowel, it is always written below the vowel. The structure of a Hangul character is shown in Fig. 1, where V1 denotes the vertically elongated vowel and V2 denotes the horizontally elongated vowel. Head and bottom consonants are denoted as C1 and C2, respectively. According to the shape of the vowel and the existence of the bottom consonants, the type of Hangul character can be divided into six categories (See Fig. 1).

III. Proposed Character Recognition system

A. The consist of proposed system

In this paper, Firstly, it gets features of mesh, projection and cross distance feature from character images. And their feature is converted into data of time series. Then using modified Henon system suggested in this paper, each character's attractor about standard Korean Character, KSC 5601 is reconstructed (See Fig. 2).

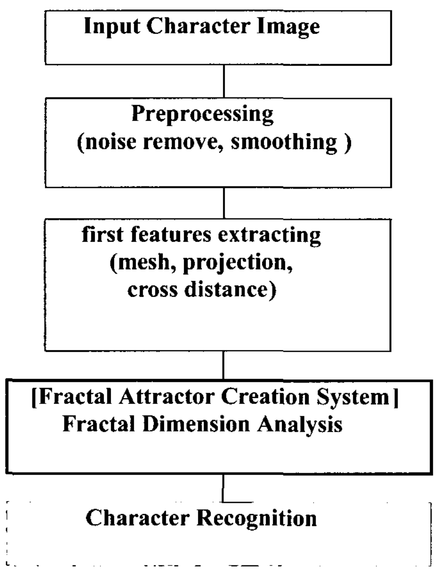


Fig. 2 Block diagram of proposed system

B. Model of the Attractor Creation

Henon's Attrator is a discrete dynamical system in two dimensions. It was suggested by the French astronomer Michel Henon in 1976 as a simplified model for the dynamics of the Lorenz system[6]. Because of its simplicity, it lends itself to computer studies and numerous investigations followed. Moreover, the gently swirling, boomerang-like shape of the attractor that arises through the dynamics is very appealing aesthetically. This object is now knows as the Henon Attractor (See Fig. 3).

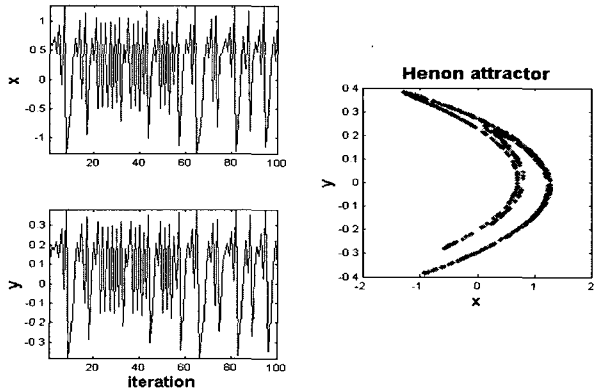


Fig. 3 The Henon Attractor

In a way which we will specify, the Henon system leads from the one-dimensional dynamics of the quadratic transformation to high-dimensional strange attractors.

The stretch-and fold action of the Henon system happens in two dimensions, with coordinates denoted by x and y. The transformation, thus, is a transformation in the plane which operates just like one of the affine transformations from our paradigm [2]. Explicitly, Henon suggested a transformation is

$$H(x, y) = (y + 1 - ax^2, bx) . \qquad (3.1)$$

where a and b are adjustable parameters, An orbit of the system consists of a starting point (x_0, y_0) and its iterated image, i.e.,

$$H(x_{k+1}, y_{k+1}) = (y_k + 1 - ax_k^2, bx_k), \quad k = 0, 1, 2, 3, \ldots \qquad (3.2)$$

Similar to the logistic equation, these dynamics depend dramatically on the choice of the constants a and b besides that of the starting point. For some parameters almost all orbits tend to a unique periodic cycle, while chaos seems to reign for other choices, Henon used the values

$$a=1.4 \text{ and } b=0.3$$

C. Creation of Character Attractor

In this paper, Firstly, it gets first feature extracting from input character images. And their feature is converted into data of time series. Then using modified Henon system suggested in this paper, character attractors is reconstructed, i.e.,

$$H'[x_k, y_k] = [y_k + 1 - a(x_k+cf_i)^2, b(x_k+cf_i)], \quad k = 0, 1, 2, 3, .. n \qquad (3.3)$$

Where cf_i is the first features of input character image. As well, in experiment, Modified Henon system parameters is adopted

$$a = 0.55 \text{ and } b = 0.3$$

In this case is reconstructed of best character attractors (see Fig. 4).

D. Analysis of Fractal Dimension

The concept of fractal dimension has inspired scientists to a host of interesting new work and fascinating speculations. Indeed, for a white it seemed as if the fractal dimensions would allow us to discover a new order in the world of complex phenomena and structures.

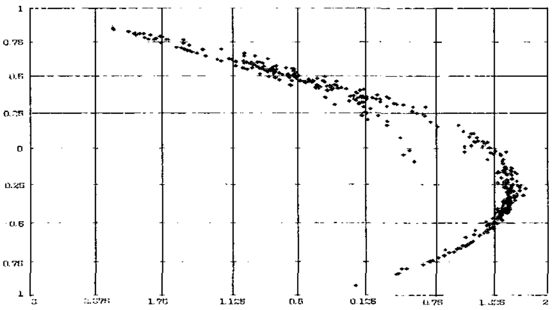


Fig. 4 Attractors of character image (Hangul '박(park)')

In this paper, in order to analyze Chaotic degree of each character's attractor, it gets last features of character image after calculating box-counting Dimension, Natural Measure, Information Bit, Information Dimension which is meant fractal dimension.

1) Box-Counting Dimension

The Box-counting dimension is related to the self-similarity dimension and the most used in measurements in all the sciences. The reason for its dominance lies in the easy and automatic computability by machine. It is straightforward to count boxes and to maintain statistics allowing dimension calculation.

We put the structure onto a regular mesh with mesh size s (Scaling factor), and simply count the number of grid boxes which contain some of the structure. This give a number, say N. Of course, this number will depend on our choice of s. Therefore we write N(s). Next we make a log/log-diagram. We then try to fit a straight line to the plotted points of the diagram and measure its slope D_f. This number is the box-counting dimension (see eq.(3.4)).

$$D_f = \lim_{s \to 0} \frac{\log N(s)}{\log 1/s} \qquad (3.4)$$

where, scaling factor s = 1/8.

2) Natural Measure

To computer Box-counting dimension we first find ectangular region that contains the whole attractor. To overcome this shortcoming of fractal dimension, boxes should be weighted according to how many times an orbit visits them. Consider an open subset *B* of a space *X* in which an attractor lies, for example, a subset *B* of the plane or the Euclidean three-dimensional space. Orbits that are typically observed in computer studies seem to eventually fill up the attractor densely.

We can count the number of times an orbit $x_0, x_1, x_2, \ldots \in X$ enters the subset *B*, and it natural to assume that the percentage of all points which are in *B* stabilizes as we perform more and more iterations. This percentage is called the natural measure $\mu(B)$ for the system, Formally.

$$\mu(B) = \lim_{n \to \infty} \frac{1}{n+1} \sum_{k=0}^{n} 1_B(x_k, y_k) \qquad (3.5)$$

where $1_B(x, y)$ is a function which is 1 or 0 when x is in *B* or not. In other words,

$$1_B(x_k, y_k) = \begin{cases} 1 & \text{if } (x_k, y_k) \in B \\ 0 & \text{otherwise} \end{cases} \qquad (3.6)$$

and $\sum_{k=0}^{n} 1_B(x_k, y_k)$ is the number of points from the orbit $x_0, x_1, x_2, \ldots x_n$ which fall in the set *B*.

The natural measure can be understood as a means of quantifying the mass of a portion of the attractor.
The calculated result of natural measure shown in Figure 5.

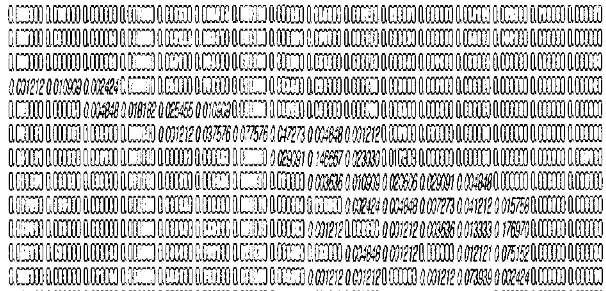


Fig. 5 Natural Measure (Hangul '박(park)')

3) Information Bit & Dimension

The theoretical foundation had been given by Claude Shannon in 1948. The logical approach would be to replace the simple box-count by a counting procedure in which each box is weighted according to its natural measure. Thus, places which the orbit passes through very frequently have a stronger impact on the calculation than boxes which the orbit rarely visit.

In terms of a formula, we replace log N(s) by

$$I(s) = \sum_{k=1}^{N(s)} \mu(B_k) \log_2 \frac{1}{\mu(B_k)} \qquad (3.7)$$

Here the sum ranges over all N(s) boxes B_k of linear size s that cover the attractor. This quantity I(s) specifies the amount of information necessary to specify a point of the attractor to within an accuracy of s; or in other words, it is the information obtained in making a measurement that is uncertain by an amount s.

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0117	0.0711	0.0211	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0373	0.1051	0.1348	0.0711	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0117	0.1779	0.2861	0.2081	0.0373	0.0117	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1485	0.4062	0.1253	0.0711	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0295	0.0711	0.1154	0.1485	0.0373	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0211	0.0373	0.0517	0.1896	0.0944	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0117	0.0000	0.0117	0.0295	0.0830	0.4421	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0373	0.0117	0.0000	0.0772	0.2806	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0117	0.0117	0.0000	0.0117	0.2778	0.0211	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0117	0.0000	0.0000	0.0295	0.0830	0.0117	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0117	0.0117	0.0446	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0446	0.0117	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0117	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Fig. 6 Information Bit (Hangul '박(park)')

The calculated result of Information bit, *I(s)* shown in Figure 6. And then the Information dimension D_I is obtained in eqn. (3.8).

$$D_I = \lim_{s \to 0} \frac{I(s)}{\log_2 1/s} \qquad (3.8)$$

IV. IMPLEMENTATION AND EXPERIMENTAL

A. Implementation

Firstly, Using modified Henon system suggested in this paper, each character's attractor about standard Korean Character, KSC 5601 is reconstructed. And then, in order to analyze Chaotic degree of each character's attractor, it gets last features of character image after calculating box-counting Dimension, Natural Measure, Information Bit, Information Dimension which is meant fractal dimension.

Even though calculation of Box-counting Dimension for character attractor in character recognition is very simple, it only determines whether attractor exists in a fixed size of box. So it is so sensitive to noise and doesn't well reflect characteristics information of character. Therefore, in this research, using information dimension and information bit computed based on Natural Measure which shows characteristic information of character, we do global classification and detailed classification by means of statistical method using minimum distance value of characteristic vector.

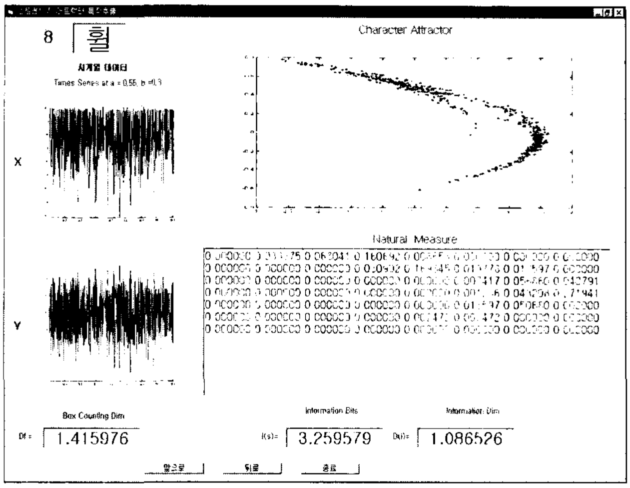


Fig. 7 Fractal Dimension at s=1/8

B. Result and Analysis

As well, in experiment, we adopt Henon system parameter (a=0.55, b=0.3) for reconstruct of best character attractor and adopt proper size of Box Count for seeking for best fractal dimension characteristic value in behalf of high precision character recognition from character attractor.

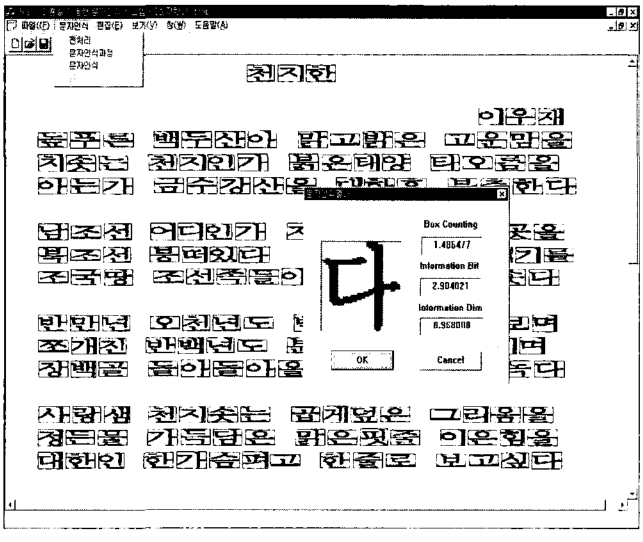


Fig. 8 Recognition Result

Also we developed a strains attractor simulator in order to discriminate Chaos status quantitatively. As we input time series data for analysis, this simulator automatically reconstructed attractor in state space and automatically calculates its dimension at high speed.

V. CONCLUSION

Chaos theory is a study researching the irregular, unpredictable behavior of deterministic and non-linear dynamical system. The interpretation using Chaos makes us evaluate characteristic existing in status space of system by time series, so that the extraction of Chaos characteristic understanding and those characteristics enables us to do high precision interpretation. Therefore, this paper proposes the new method which is adopted in extracting character features and recognizing character using fractal dimension of Chaos theory which highly recognizes a minute difference with strange attractor created from Modified Henon system. This papers implements a high precision character recognition system. The experimental result shows 99% character classification rates for 2,350 Korean characters using proposed method in this papers.

REFERENCES

[1] Robert R. Trippi, "Chaos & Nonlinear Dynamics in the Financial Markets", IRWIN, 1995.

[2] H. O. Peitgen, H. Jurgens, D. Saupe, "Chaos and Fractals: New Frontiers of Science", Springer-Verlag, 1992.

[3] Edward Ott, "Chaos in Dynamical Systems", Cambridge, 1993.

[4] S. H. Kellert, "In the Wake of Chaos : Unpredicatable Order In Dynamical System", univ. of Chicago Press, 1993.

[5] James Gleick, "CHAOS: Making a New Science", 1993.

[6] M. Henon, "A two-dimensional mapping with a strange attractor", Comm. Math. Phys. 50, pp. 69-77, 1976.

[7] M. Nadler and E. P. Smith, "Pattern Recognition Engineering", New York, 1993.

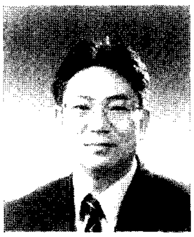
[8] M. Nadler and E. P. Smith, "Pattern Recognition Engineering", New York, 1993.

[9] K. S. Fu, "Syntactic Methods in Pattern Recognition", academic press, New York, 1970.

[10] Y. W. Shon and J. C. Namkung, "A study on the Hangeul confusion character recognition using fractal dimension and attractor," KIPS, vol. 6, no 7, pp. 1825-1831, 1999.

[11] Y. W. Shon, "A study on the confusion character recognition using the modified Henon attractor," IST-2003 International Conference, NSTU, Russia, tom 3. pp. 103-109, 2003.

[12] Y. W. Shon, "Printed Numeric character recognition using Fractal Dimension and modified henon Attractor,"KMS, vol.6, no.1, pp.89-96. 2003.



Young-Woo Shon

He received his B.S. and M.S. degree from the department of Electronic Engineering, Kwangwoon University in 1981 and 1983, respectively. He received his Ph.D. degree from the department of Computer Engineering, Kwangwoon University in 2000. He worked as a senior research in KINITI (Korea INstitute of Industry & Technology Information), Seoul Korea from 1991 to 1997. He is current an assistant professor of Computer Science at Kimpo College, Korea. His research interests include Chaos Theory, Fractal and Attractor Analysis, Image Processing, Pattern Recognition, Web application of Multimedia.