

# Wavelet-Based Fuzzy Modeling Using a DNA Coding Method

Young-Hoon Joo\* and Yeun-Woo Lee\*

\*School of Electronic & Information Eng, Kunsan National University, Chonbuk, Korea

## Abstract

In this paper, we propose a new wavelet-based fuzzy modeling using a DNA coding method. Generally, it is well known that the DNA coding method is more diverse in the knowledge expression and better in the optimization performance than the genetic algorithm (GA) because it can encode more plentiful genetic informations based on the biological DNA. The proposed method can construct a fuzzy model using the wavelet transform, in which the coefficients are identified by the DNA coding method. Thus, we can effectively get the fuzzy model of the nonlinear system by using the advantages of both wavelet transform and DNA coding method. In order to demonstrate the superiority of the proposed method, it is compared with modeling method using the conventional GA.

**Key words :** DNA coding, fuzzy modeling, genetic algorithm, wavelet transform

## 1. Introduction

Fuzzy model is more effective in representing the human mind and natural language, and more suitable for dealing with the complex, nonlinear, ill-defined and uncertain system than the conventional mathematical models. In realistic systems, we can hardly obtain the accurate differential equation since there are so many nonlinearities. Therefore, several fuzzy modeling methods, which can represent nonlinear systems, have been proposed in recent years.

Fuzzy logic theory was introduced at first by Zadeh [1] and gave a new possibility in each field. Tong [2] proposed a new fuzzy-logic-based approach in modeling of complex nonlinear system. After that, Pedrycz [3] proposed a new identification algorithm for a fuzzy system based on conditional probability distribution, and Xu [4] proposed an identification method for a general fuzzy system, which included the structure identification of fuzzy rule, parameter estimation, and self-learning algorithm. Also, Sugeno [5] proposed a general identification method which could obtain qualitative model by using only input/output data of nonlinear system and replaced the consequent part of the fuzzy rule by a linear equation. As stated above, the main purpose of these efforts is to describe the unknown system as a set of fuzzy rules as correctly as possible.

To identify the parameters and the structure of a fuzzy model, the genetic algorithm (GA) has been widely applied for various nonlinear system. But the conventional GA encoded the solution space into the string with fixed position and the fixed length, and to acquire the proper linkages associated with a given problem was difficult because the structure of encoded solution was unknown. So, this weak linkage means that the building-block is likely to breakdown. To prevent this problem, Joo proposed a hybrid algorithm using both the GA and the fuzzy c-means clustering [6], and

a fuzzy modeling using the messy genetic algorithm (mGA) [7]. And recently, the DNA coding method was studied to solve the problem of GA [8].

The DNA coding method is more diverse in the knowledge expression and better in the optimization performance than the GA because it can encode more plentiful genetic information based on the biological DNA [8]. A living body starts its life from the DNA, the primary information carrier in genetics. Almost every critical activity of the organism is accomplished by proteins constructed from the DNA. The efficacy of the organism, i.e, the genetic fitness, depends on the proteins. For some reason our body chooses different representations of the information stored in the proteins. It uses the mRNA and the DNA sequences to represent the proteins.

On the other hand, Wang and Zeng [9] represented the fuzzy system with the linear combination of the fuzzy basis function (FBF), and Lin [10] obtained the equivalent model to the discrete wavelet transform through the modification of the fuzzy model. Hereby, the problem that the conventional fuzzy model can hardly deal with the abrupt change of signal was resolved. The wavelet transform was very effective in analyzing the physical status of a certain signal with singularity and with an element of high frequency as compared with the Fourier transform. The wavelet theory has been developed by Daubechies [12], and made rapid progress by proof of Donoho [11] that wavelet can be a basis function for any signal using unconditional basis. The fuzzy system modeling based on Donoho's unconditional basis has the advantage of wavelet transform by constituting FBF and consequent part to equalize the linear combination of FBF with the linear combination of wavelet function and modifying fuzzy system model to be equivalent to wavelet transform. The conventional fuzzy model may not give the satisfactory result to transient signal. On the contrary, in the wavelet-based fuzzy system model, the accurate fuzzy model can be obtained because energy compaction by the unconditional basis and the description of transient signal by wavelet basis functions is distinguished.

In this paper, we propose a new wavelet-based fuzzy

modeling using a DNA coding method. Modifying conventional fuzzy model to have several fuzzy rule bases and the consequent part to be the part of the 'Mexican Hat' wavelet function, the fuzzy model was equivalent to discrete wavelet function. The wavelet-based fuzzy model, thus, inherited the advantage from discrete wavelet transform. When an arbitrary signal is represented with linear combinations of wavelet functions, the energy can be compacted.

Also, the DNA coding method is utilized to identify the parameters and rule number of fuzzy system. According to genetic information of amino acid, fuzzy rule is created. And, each individual evolve using the genetic operators - crossover, deletion, insertion, and inversion. Individual with high fitness is reproduced in next generation via these process.

Finally, proposed method is applied to fuzzy modelling of Mackey-Glass chaos time series system, that is the representative nonlinear system, in order to demonstrate the superiority of the proposed method compared with the GA.

## 2. Discrete Wavelet Transform and Fuzzy Model

### 2.1 Discrete Wavelet Transform

Any function or signal in  $L^2(R)$  can be represented as the linear combination of basis function in the following form[17];

$$f(x) = \sum_k c_{j,k}(k) \varphi_{j,k}(x) + \sum_k \sum_{j=j_0} d_j(k) \Psi_{j,k}(x) \quad (1)$$

where,  $\varphi_{j,k}(x)$  is scale function and  $\Psi_{j,k}(x)$  is the wavelet function. In Eq. (1), the first sum is approximation of low resolution and the second sum is approximation of high resolution. The wavelet transform is the mapping of  $L^2(R) \rightarrow L^2(R^2)$  and is separated into a low band pass filter, scale function and a high band pass filter, wavelet function. If  $\varphi_{j,k}(x)$  and  $\Psi_{j,k}(x)$  is orthogonal each other for any  $j$  and  $k$  the coefficients of Eq.(1)  $c_j(k)$  and  $d_j(k)$  can be obtained by inner product. Finding the coefficients is called wavelet transform. The multi-resolution equation can represent the signal with time-scale by dividing into precise elements. Therefore,  $\Psi_{j,k}(x)$  of Eq.(1) can be represented by,

$$\Psi_{j,k}(x) = a^{-j/2} \Psi(a^{-j}x - bk) \quad (2)$$

A multidimensional wavelet function is represented with tensor product of single dimensional wavelet function as follows;

$$\Psi(x) = \Psi_1(x_1) \cdots \Psi_n(x_n) \quad (3)$$

Assuming that single dimensional wavelet transform is separated into  $n$  orthogonal direction elements, Fourier transform of each term in Eq.(3) is substituted for itself.

$$\hat{\Psi}(\omega) = \hat{\Psi}_1(\omega_1) \cdots \hat{\Psi}_n(\omega_n) \quad (4)$$

where,  $\hat{\Psi}(\omega)$  is Fourier transform of  $\Psi(x)$ .

$$\int \frac{|\hat{\Psi}_i(\omega_i)|^2}{|\omega_i|} d\omega_i < \infty \quad (5)$$

It is proven that admissibility condition of Eq.(5) must be satisfied. On condition of attenuation, Eq.(5) can satisfy the following for  $\Psi_i(x_i)$  that converges into 0 for  $+\infty$  and  $-\infty$ .

$$\int \Psi_i(x_i) dx_i = 0 \quad (6)$$

As proven in [17], the following equation should be satisfied to be used as wavelet frame.

$$A \|f\|^2 \leq \sum_{j,k} |(f, \Psi_{j,k})|^2 \leq B \|f\|^2 \quad (7)$$

where,  $A > 0, 0 < B < \infty$ . Therefore,  $\Psi_i(x_i)$  which satisfies conditions Eq.(6) and Eq.(7) should be set as the wavelet frame. In this thesis, 'Mexican Hat' is employed as a mother wavelet function which satisfies both of the conditions as follows;

$$\Psi_i(x_i) = \alpha_i (1 - \alpha_i x_i^2) e^{-\frac{\alpha_i x_i^2}{2}} \quad (8)$$

'Mexican Hat' in Eq.(8) is derived from the function which is in proportion to the second-ordered differential form of the Gaussian probability density function.

Substitute Eq.(8) for Eqs.(2) and (3) yields the followings;

$$\Psi_{j,k} \approx a^{\frac{j_1}{2}} \alpha_1 [1 - \alpha_1 (a^{-j_1} x_1 - b_1 k_1)^2] e^{-\frac{\alpha_1 (a^{-j_1} x_1 - b_1 k_1)^2}{2}} \cdots - a^{\frac{j_n}{2}} \alpha_n [1 - \alpha_n (a^{-j_n} x_n - b_n k_n)^2] e^{-\frac{\alpha_n (a^{-j_n} x_n - b_n k_n)^2}{2}} \quad (9)$$

### 2.2 Fuzzy model

The fuzzy model is modified so that it can be equivalent to discrete wavelet transform.

The fuzzy rule is to represent the characteristic of a system linguistically. The fuzzy model which is used in this thesis is as follows;

$$\begin{aligned} \text{Rule } i : & \text{ If } x_1 \text{ is } A_{i1}, \dots, x_n \text{ is } A_{in}, \\ & \text{ Then } y_i \text{ is } d_i \alpha_1 (1 - \alpha_1 x_1^2) \alpha_2 (1 - \alpha_2 x_2^2) \\ & \quad \cdots \alpha_n (1 - \alpha_n x_n^2) \end{aligned} \quad (10)$$

where, Rule  $i$  is the  $i$ -th rule,  $x_j$  is the  $j$ -th input variable,  $y_i$  is the  $i$ -th output variable and  $A_{ij}$  is a membership function for the  $i$ -th rule of the  $j$ -th input defined as a Gaussian function. The consequent part constituted by the product of the remainders except Gaussian function of 'Mexican Hat' wavelet function in Eq.(9).

That is, the model which is equivalent to wavelet transform can be obtained by modifying the consequent part of the general fuzzy model.

The output for the arbitrary fuzzy rule basis in the fuzzy inference system which has several fuzzy rule bases can be

represented with the linear combination of fuzzy basis function(FBFs) in the following;

$$y_j = \sum_{i=1}^n B_i(x)y_i \quad (11)$$

Substitute Eq. (10) for Eq. (11) yields the followings;

$$y_j = \sum_{i=1}^n B_i d_i a^{\frac{j_i}{2}} \alpha_1 (1 - \alpha_1 (a^{-j_i} x_1 - b_1 k_1)^2) \dots a^{\frac{j_n}{2}} \alpha_n (1 - \alpha_n (a^{-j_n} x_n - b_n k_n)^2) \quad (12)$$

where,  $B_i = \frac{\prod_{k=1}^n A_{ik}(x_k)}{\sum_{i=1}^n \prod_{k=1}^n A_{ik}(x_k)}$ , and  $d_i$  is a real number.

$$y = \sum_{j=1}^N y_j = \sum_{j=1}^N \sum_{k=1}^n d_{j,k} \Psi_{j,k} \quad (13)$$

Eq. (13) means that the fuzzy model is equivalent to discrete wavelet transform through modifying the consequent part of a fuzzy model which has several fuzzy rule bases. Using 'Mexican hat' wavelet function, the Gaussian function part in the wavelet function becomes membership function of the fuzzy model,  $A_{ik(x)}$  and the product of the parts except the Gaussian function becomes the consequent part of the modified fuzzy model.

### 3. Identification of fuzzy model using a DNA coding method

The DNA coding method is optimization technique based on biological DNA structure. It is more diverse in the knowledge expression and better in the optimization performance than the GA because it can encode more plentiful genetic information.

The DNA is the primary carrier of the genetic information that is transmitted from one generation to another, and DNA molecules consist of two long complementary chains held together by base pairs. DNA consists of four kinds of bases joined to a sugar-phosphate backbone. The four bases in DNA are Adenine (A), Guanine (G), Thymine (T) and Cytosine (C). Chromosomes are made of DNA double helices. Where, bases in DNA helices obey the complementary base pairing rule. T and G pair with A and C respectively.

The information coded in the DNA is extracted during the process of gene expression and expression of genetic information coded in DNA requires construction of the mRNA sequence, followed by that of proteins. Also, DNA is decoded by mRNA and is translated by protein in ribosome. The main steps are,

- *transcription* : formation of mRNA from DNA
- *translation* : formation of protein from mRNA
- *protein folding*

Figure 1 shows the above process. Basic unit of biological DNA is the four nucleotide. Nucleotide is consisted of A, G,

C, and T. It is comprised of three adjacent nucleotides in DNA and forms one codon. Such 64 codons are translated by amino acid which have genetic information.

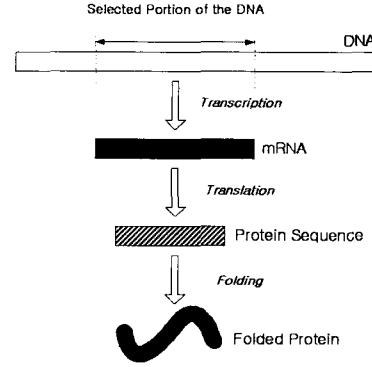


Fig. 1 Expression of genetic information coded

Value in parenthesis that display the genetic information of each amino acid play the role that specify the voluntarily segmentalized value in input and output studying data in Table.1. Table.1 shows the translation from codons into amino acids.

Table 1. Translation from codons to amino acids

1st position (5'end)	2nd position(middle)				3rd position (3'end)
	T	C	A	G	
T	Phe(1)	Ser(10)	Tyr(4)	Cys(19)	T
	Phe(1)	Ser(10)	Tyr(4)	Cys(19)	C
	Leu(9)	Ser(10)	Stop(0)	Stop(0)	A
	Leu(9)	Ser(10)	Stop(0)	Trp(20)	G
C	Leu(9)	Pro(7)	His(5)	Arg(11)	T
	Leu(9)	Pro(7)	His(5)	Arg(11)	C
	Leu(9)	Pro(7)	Gln(14)	Arg(11)	A
	Leu(9)	Pro(7)	Gln(14)	Arg(11)	G
A	Ile(2)	Thr(8)	Asn(15)	Ser(10)	T
	Ile(2)	Thr(8)	Asn(15)	Ser(10)	C
	Ile(2)	Thr(8)	Lys(16)	Arg(11)	A
	Met(3)	Thr(8)	Lys(16)	Arg(11)	G
G	Val(6)	Ala(12)	Asp(17)	Gly(13)	T
	Val(6)	Ala(12)	Asp(17)	Gly(13)	C
	Val(6)	Ala(12)	Glu(18)	Gly(13)	A
	Val(6)	Ala(12)	Glu(18)	Gly(13)	G

Therefore, codon is translated into amino acid. According to genetic information of amino acid, fuzzy rule is created. Also, each individual evolve using the genetic operators - crossover, deletion, insertion, and inversion. Individual with high fitness is reproduced in next generation via these process.

Fig. 2 shows arithmetic principle of each genetic operator, and Fig. 3 shows genetic operators proposed in this paper,

Since the GA is guided only by the fitness function, we have to consider the better fitness function to determine the accuracy and the size of the fuzzy inference system. The performance measures of the accuracy and size of the fuzzy inference system is as follows;

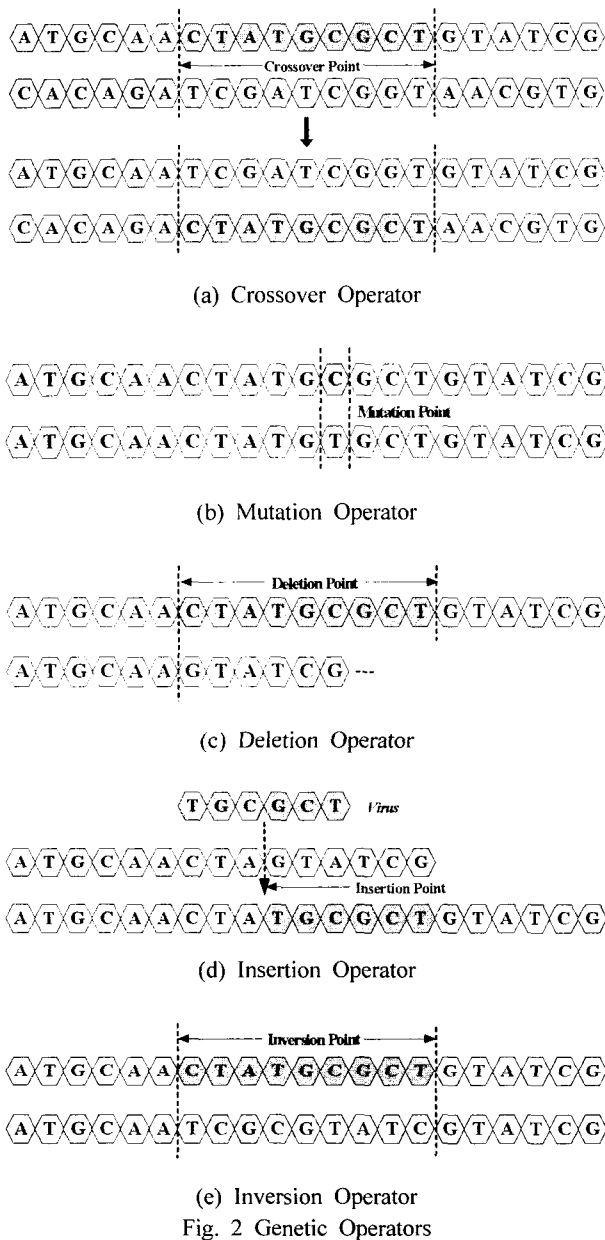


Fig. 2 Genetic Operators

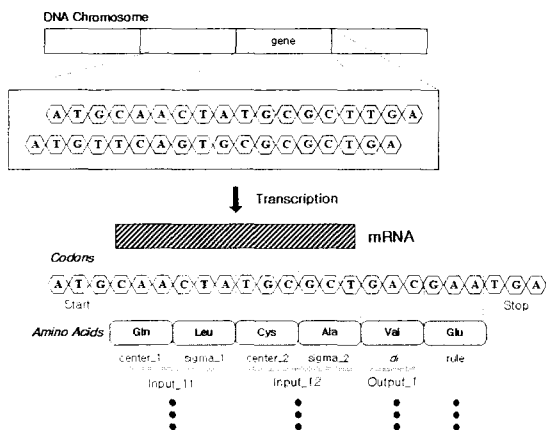


Fig. 3 Generation of Fuzzy Rules

$$f(PI_{accuracy}, PI_{size}) = \lambda \frac{1}{1 + PI_{accuracy}} + (1 - \lambda) \frac{1}{1 + PI_{size}} \quad (14)$$

where,  $\lambda$  ( $0 \leq \lambda \leq 1$ ) is the weighting factor. Fuzzy model with null set cannot infer suitable outputs from the given input data set. In order to solve this problem, we introduce the following type of penalty function for the fitness function;

$$P = \sum_{i=1}^N p_i \quad (15)$$

where,  $p_i$  is a constant value if there exists null set in the fuzzy inference system for the given  $i$ -th input data set, otherwise  $p_i$  is zero.

Also, by combining Eq.(14) and Eq.(15), we have obtain the following fitness function;

$$f(PI_{mse}, PI_{rule}) = \left\{ \lambda \frac{1}{1 + PI_{mse}} + (1 - \lambda) \frac{1}{1 + PI_{rule}} \right\} \frac{1}{1 + P} \quad (16)$$

#### 4. Simulation

To demonstrate the superiority of the proposed fuzzy modeling method, chaotic time series data which was presented by the Mackey-Glass is used [14]. This data is obtained as follows:

$$\dot{x}(t) = \frac{0.2x(t-\tau)}{1 + x^{10}(t-\tau)} - 0.1x(t) \quad (17)$$

The goal of this task is to use past values of the time series up to time  $t$  to predict the value at some point in the future  $t + P$ . The standard method for this type of prediction is to create a mapping from  $D$  points of the time series spaced  $\Delta$  apart, that is,  $[x(t - (D - 1)\Delta), \dots, x(t - \Delta), x(t)]$ , to a predicted future value  $x(t + P)$ . To allow comparison with earlier work, the values  $D = 4$  and  $\Delta = P = 6$  were used. We applied the fourth-order Runge-Kutta method to find the numerical solution. The time step used in the method was 0.1, initial condition  $x(0) = 1.2$ , and  $\tau = 17$ . In this way,  $x(t)$  was thus obtained via numerical integration for  $25 \leq t \leq 524$ . From the Mackey-Glass time series  $x(t)$ , we extracted 500 input-output data pairs of the following format[14];

$$[x(t-18), x(t-12), x(t-6), x(t); x(t+6)] \quad (18)$$

Table.2 shows initial parameters of the GA and the DNA coding method.

Table 2. Initial parameters

Parameters	Value
Generation (DNA, GA)	200
Population (DNA, GA)	300
Crossover rate (DNA, GA)	0.9
Mutation rate (DNA, GA)	0.01
Deletion rate (DNA)	
Insertion rate (DNA)	
Inversion rate (DNA)	

Table 3. Comparisons between our model and other models

Method	Non-dimensional error index
ANFIS[14]	0.007
AR model[14]	0.19
Cascaded-correlation NN[15]	0.06
Backpropagation MLP[15]	0.02
6th-order polynomial[15]	0.04
Linear predictive method[15]	0.55
Ours	0.001

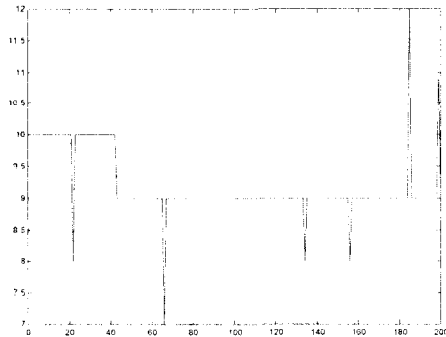


Fig. 4 Change of rules

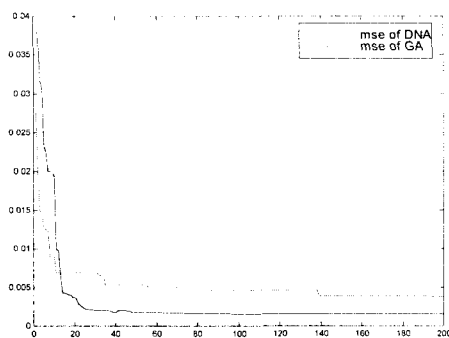


Fig. 5 Change of RMSE for the GA and the DNA coding method

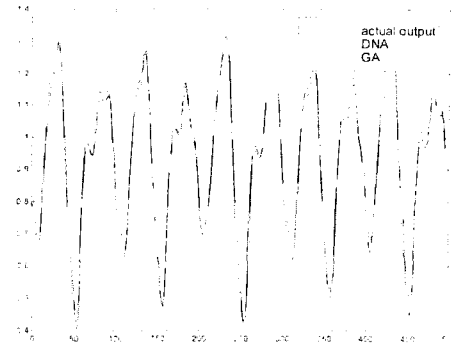


Fig. 6 The identification result

Simulation results are shown in Fig.4, 5 and 6. Figure 4 shows the change of number of rules, and Fig. 5 shows the changes of RMSE for the GA and the DNA coding method. Fig.5 shows the comparison among the actual output and the model outputs for the proposed method and the GA. It can be known that the proposed method is robust to chaotic phenomenon. As shown in Table.3, the non-dimensional error index (NDEI) of the proposed method is superior to those of other methods. Thus, it is proven that conventional fuzzy system models have the incorrect output in prediction but the wavelet-based fuzzy system model is well predictive on the new input data. In particular, there was not much difference between the model prediction and the model output. It could be visually confirmed that the proposed method is very robust.

## 5. Conclusion

In this paper, we proposed a new wavelet-based fuzzy modeling using a DNA coding method. Modifying conventional fuzzy model to have several fuzzy rule bases and the consequent part to be the part of the 'Mexican Hat' wavelet function, the fuzzy model was equivalent to discrete wavelet function. The wavelet-based fuzzy model, thus, inherited the advantage from discrete wavelet transform. When an arbitrary signal is represented with linear combinations of wavelet functions, the energy can be compacted. It was confirmed that the prominent fuzzy model can be obtained with the small number of rules by this property. In order to, demonstrate the superiority and efficiency of the proposed scheme, we applied this method to the fuzzy modeling of the 'Mackey-Glass' chaotic time series system.

## References

- [1] L. A. Zadeh, "Fuzzy Sets", *Int. Jour. of Information control*, vol. 8, pp. 338-353, 1965.
- [2] R. M. Tong, "The Evaluation of Fuzzy Models Derived from Experimental Data", *Fuzzy Sets and Systems*, vol. 4, pp. 1-12, 1980.
- [3] W. Pedrycz, "An Identification Algorithm in Fuzzy

Relational Systems", *Fuzzy Sets and Systems*, vol. 13, pp. 153-167, 1984.

- [4] C. W. Xu, "Fuzzy Model Identification and Self-learning for Dynamic Systems", *IEEE Trans. Sys. Man. Cybern.*, vol. 17, No. 4, pp. 683-689, 1987
- [5] M. Sugeno and T. Yasukawa, "A Fuzzy Logic Based Approach to Qualitative Modeling", *IEEE Trans. Fuzzy Sys.*, vol. 1, pp. 7-31, 1993.
- [6] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Fuzzy System Modeling by Fuzzy Partition and GA Hybrid Schemes", *Fuzzy Sets and Systems*, vol. 86, No. 3, pp. 279-288, 1997.
- [7] Y. H. Joo, Y. W. Lee, J. B. Park, "mGA Based Fuzzy Model Identification", *Electronics Letters*, vol. 49D, No. 8, pp. 407-414, 2000. 8.
- [8] T. Yoshikawa, T. Furuhashi, Y. Uchikawa, "A Combination of DNA Coding Method with Pseudo-Bacterial GA for Acquisition of Fuzzy Control Rules", *The 1st Online Workshop on Soft Computing(WCSI)*, 1996. 8
- [9] X. J. Zeng and M. G. Singh, "Approximation Accuracy Analysis of Fuzzy Systems with the Center-Average Defuzzifier", *Proc. FUZZ-IEEE/IFES'95*, pp. 109-116, 1995.
- [10] C. K. Lin and S. D. Wang, "Fuzzy Modeling Using Wavelet Transforms", *Electronics Letter*, vol. 32, pp. 2255-2256, 1996.
- [11] A. Bruce, D. Donoho, H. Y. Gao, "Wavelet Analysis for Signal Processing", *IEEE Spectrum*, vol, 33, pp. 26-35, 1996
- [12] I. Daubechies, "The Wavelet Transform, Time-frequency Localization and Signal analysis," *IEEE Trans. on Inform. Theory*, Vol. 36, pp. 961-1005, 1990.
- [13] M. C. Mackey and L. Glass. "Oscillation and Chaos in Physiological Control Systems", *Science*, vol. 197, pp. 287-289, 1977.
- [14] J. S. Jang. "ANFIS: Adaptive-Network-based Fuzzy Inference Systems", *IEEE Transactions on Systems, Man, and Cybernetics*, 23(03):665-685, 1993.
- [15] A. S. Lapedes and R. Farber. "Nonlinear signal processing using neural networks: prediction and system modeling", *Technical Report LA-UR-87-2662*, 1987.
- [16] R. S. Crowder III, D. Touretzky, G. Hinton and T. Sejnowski. "Predicting the Mackey-Glass time-series with cascade-correlation learning". *Proceedings of the 1990 Connectionist Models Summer School*, pp. 117-123.
- [17] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, 1996.



**Young-Hoon Joo**

He received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Yonsei University, Korea, in 1982, 1984, and 1995, respectively. He worked with Samsung Electronics Company, Korea, from 1986 to 1995, as a Project Manager. He was with University of Houston, TX, from 1998 to

1999, as a Visiting Professor in the Department of Electrical and Computer Engineering. He is currently Associate Professor in the School of Electronic and Information Engineering, Kunsan National University, Korea. His major is mainly in the field of mobile robots, fuzzy modeling and control, genetic algorithm, intelligent control, and nonlinear systems control. Prof. Joo is now serving as the Associate Editor for the *Transactions of the Korea Institute of Electrical Engineers* and Editor-in-Chief for *Korea Journal of Fuzzy Logic and Intelligent Systems*.

Phone : +82-63-469-4706, Fax : +82-63-469-4706

Email : yhjoo@kunsan.ac.kr



**Yeun-Woo Lee**

He received the B. S. and M. S. degrees in the department of control & instrumentation engineering from Kunsan National University, Korea in 1998 and 2000, respectively. He is currently working toward the Ph.D. degree in the department of Electronic and Information Engineering,

Kunsan National University, Korea. His current research interests are in fuzzy modeling, genetic algorithms, DNA computing, and intelligent control.

Phone : +82-63-469-4706 Fax : +82-63-469-4706

E-mail : raic@kunsan.ac.kr