An Improved Entropy Function and Chaos Optimization Based Scheme for Two-Dimensional Entropic Image Segmentation

Cheng Ma and Chengshun Jiang

Institute of Information Engineering
Information Engineering University, Zhengzhou, 450002, China
macheng31@gmail.com

Abstract. An improved two-dimensional entropic image segmentation method is presented in this paper. The method makes use of a new entropy function defined in a simple form, which can reduce computational amount notably. And the correctness of the new function is also proved. Then a scheme based on mutative scale chaos optimization is adopted to search for the optimal threshold. The results of simulation illustrate that efficiency of segmentation is improved significantly due to the new entropy function and searching method.

1 Introduction

The thresholding method based on maximum entropy is one of the most widely used methods in image segmentation. It uses the gray-level features of an image to choose a single or multiple thresholds by which the image pixels are classified into several regions and thus the object is extracted from the background. The one-dimensional entropic method was firstly introduced by Kapur in 1985 [1]. Abutaleb expanded it to the two-dimensional space in 1989 [2]. Compared to 1-D method, the 2-D method makes use of pixels' gray levels and average gray levels within a neighborhood, which can produce a better segmentation result. And it also shows a stronger ability of resisting noises. However, the computational amount in 2-d method increases sharply than in 1-D situation. To solve the problem, people use some optimization methods such as genetic algorithms (GA) [3,4,5], chaos optimization method [6], etc. Some researchers focused on the simplification of the mathematical expression of the entropy function and put forward some fast algorithms [7,8]. Yang proposed a segmentation method based on an optimized entropy function which reduced the computational amount efficiently, see [9]. However we find there's some irrationality in the author's work. In this paper, we analyze the function and make some modification to it. Later we will make experiment to test the new function,based on a chaos optimization scheme.We also compare the efficiency of genetic algorithm and chaos optimization method.

2 Model of 2-D Maximum Entropic Segmentation

2.1 Conventional Model

Suppose the gray level interval of a $M \times N$ sized image is $[0, L]$, so the pixels' average gray level within a neighborhood is also in $[0, L]$. Let $f(x, y)$ denote the gray level

Y. Wang, Y. Cheung, and H. Liu (Eds.): CIS 2006, LNAI 4456, pp. 991–1000, 2007.

of pixel (x, y) and $g(x, y)$ denote the average gray level of a $n \times n$ neighborhood. Then the 2-D gray level pair of pixel (x, y) is denoted by $[f(x, y), g(x, y)]$. Let $p_{ij} = r_{ij}/(MN)$represent the probability of the gray-level pair (i, j), $(i, j = 0, 1, \cdots L-1)$ where r_{ij} is the number of elements in the event $\{f(i, j) = r_{ij}\}$. Obviously we have $0 \leq r_{ij} \leq M \cdot N$ and $\sum_{i=0}^{L-1} \sum_{i=0}^{L-1} p_{ij} = 1$. For each class of pixels labeled by A and B, let

$$H_A = -\sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij}^A \log_2 p_{ij}^A,$$

$$H_B = -\sum_{i=s}^{L-1} \sum_{j=t}^{L-1} p_{ij}^B \log_2 p_{ij}^B,$$

where

$$p_{ij}^A = p_{ij} / \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij}, \quad p_{ij}^B = p_{ij} / \sum_{i=s}^{L-1} \sum_{j=t}^{L-1} p_{ij}.$$

p_{ij}^A, p_{ij}^B denote the probability of the gray-level pair (i, j) in region A and B respectively. Then the sum of the two entropies

$$H_1(s,t) = H_A + H_B = -\sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij}^A \log_2 p_{ij}^A - \sum_{i=s}^{L-1} \sum_{j=t}^{L-1} p_{ij}^B \log_2 p_{ij}^B \tag{1}$$

is the object function. The goal of the segmentation is to find a threshold (s^*, t^*) satisfying the following nonlinear problem:

$$F_1(s^*, t^*) = \max_{0 \leq s,t \leq L-1} H_1(s,t). \tag{2}$$

2.2 A New Entropy Function

For the purpose of reducing computational amount, researchers are focusing on proposing some fast algorithms by deducing equation (1). According to the property that an entropy function reaches its maximum in an equiprobable distribution, a new entropy function was presented by Yang in [9]:

$$H_2(s,t) = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} \left| p_{ij}^A - \frac{1}{st} \right| + \sum_{i=s}^{L-1} \sum_{j=t}^{L-1} \left| p_{ij}^B - \frac{1}{(L-s)(L-t)} \right|. \tag{3}$$

and the following problem

$$F_2(s^*,t^*) = \min_{0\le s,t\le L-1} H_2(s,t) \tag{4}$$

was proven to be equivalent to problem (2). That means all logarithmic and multiplicative operations are substituted by additions. This is an effective method in reducing computational amount.

To prove the equivalence, Yang insisted that when entropy function H_1 reached its maximum, the probability p_{ij}^A in (1) equalled to the same value $1/st$ for different (i,j) while p_{ij}^B equalled to $1/[(L-s)(L-t)]$. So the new entropy function $H_2(s,t)$ in (3) reached its minimum 0. On the contrary, when $H_2(s,t)$ equalled to 0, we had $p_{ij}^A = 1/(st)$ and $p_{ij}^B = 1/[(L-s)(L-t)]$ for each (i,j). That meant entropy H_A and H_B reached their maximums in such an equiprobable distribution. Therefore the two problems were equivalent. However, some point was ignored by Yang. If p_{ij}^A and p_{ij}^B could both reach the point $1/(st)$ or $1/[(L-s)(L-t)]$ for every (i,j), the proof was correct in theory. But we know that in a given gray-scale image, gray level pairs (i,j) are mainly distributed in regions representing object and background, and some pairs don't exist. In such a situation, whatever s and t equal to, p_{ij}^A or p_{ij}^B will equal to 0, i.e. they cannot equal to $1/(st)$ or $1/[(L-s)(L-t)]$, hence $H_1(s,t)$ and $H_2(s,t)$cannot reach their extrema. For two different functions who can reach their extrema at the same point in $[0,1]$, they may have different extremum points in a smaller interval $[a,b]$, $(0\le a\le b\le 1)$. Therefore, the proof by Yang is not totally correct.

Despite of the mistake, the work Yang has done is valuable. In this paper we make some modifications on the basis of (3) and present an improved entropy function, whose correctness will be proven in the following theorem.

Theorem. Define a new object function

$$\begin{aligned} H_3(s,t) = &\sum_{i=0}^{s-1}\sum_{j=0}^{t-1}\left(p_{ij}^A - \frac{1}{st}\right)^2 \\ &+ \sum_{i=s}^{L-1}\sum_{j=t}^{L-1}\left(p_{ij}^B - \frac{1}{(L-s)(L-t)}\right)^2, \end{aligned} \tag{5}$$

where $p_{ij}^A, p_{ij}^B \in [a_{ij}, b_{ij}] \subset [0,1]$. Then the new nonlinear optimization problem

$$F_3(s^*,t^*) = \min_{0\le s,t\le L-1} H_3(s,t) \tag{6}$$

is equivalent to (2).

Proof. Denote both p_{ij}^A and p_{ij}^B by $p_{ij}(0\le i,j\le L-1)$. H_1 and H_3 can be regarded as functions of p_{ij}. For those $p_{ij}=0$ terms in H_1, we ignore them. Now let's check the existence of H_1's extremum.

Because

$$\frac{\partial(-H_1)}{\partial p_{ij}} = \log_2 p_{ij} + 1, \quad \frac{\partial^2(-H_1)}{\partial p_{ij}^2} = \frac{1}{p_{ij}} > 0$$

and for any $(i', j') \neq (i, j)$, we have

$$\frac{\partial^2(-H_1)}{\partial p_{ij}\partial p_{i'j'}} = \frac{\partial}{\partial p_{i'j'}}(\log_2 p_{ij} + 1) = 0 \ .$$

So the Hessian matrix of $-H_1(p_{ij})$ is a positive definite symmetric matrix whose diagonal elements are $1/p_{ij} > 0$ while remainders are zeros. Therefore, $-H_1$ is a strictly convex function. According to the optimization theory, $-H_1$ has a unique minimum. That means H_1 has a unique maximum point. To prove the equivalence of the two optimization problems, let's firstly check a function with two variables

$$f_1 = -\sum_{i=1}^{2} p_i \log_2 p_i, \text{where } p_1 + p_2 = 1 \ . \tag{7}$$

From (7) we have $f_1 = -p_1 \log_2 p_1 - (1-p_1)\log_2(1-p_1)$. This function has a unique maximum, as shown in Fig.1. If $p_1 = 1/2$, then f_1 reaches its maximum. However, if p_1 is limited within a smaller interval $[a, b]$ satisfying $1/2 \notin [a, b]$, then we know from the figure that the smaller the distance between p_1 and $1/2$, the bigger f_1 evaluates. That means f_1 is approaching its maximum as $(p_1 - 1/2)^2$ is approaching its minimum. When $(p_1 - 1/2)^2$ reaches its minimum in $[a, b]$,

$$\begin{aligned} 2\left(p_1 - \frac{1}{2}\right)^2 &= \left(p_1 - \frac{1}{2}\right)^2 + \left(1 - p_2 - \frac{1}{2}\right)^2 \\ &= \sum_{i=1}^{2}\left(p_i - \frac{1}{2}\right)^2 \end{aligned}$$

also reaches its minimum. So f_1 evaluating its maximum equals to $\sum_{i=1}^{2}(p_i - 1/2)^2$ evaluating its minimum.

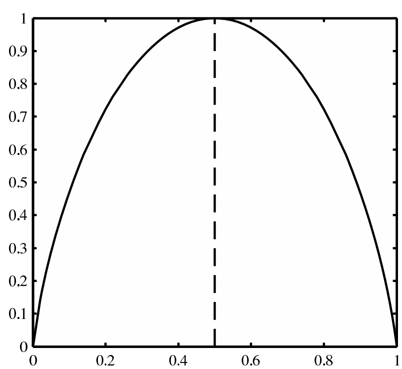

Fig. 1. Entropy function with one variable. It's a convex function with one maximum.

Let's move on to a function with three variables

$$f_2 = -\sum\nolimits_{i=1}^{3} p_i \log_2 p_i, \text{ where } p_1 + p_2 + p_3 = 1, \tag{8}$$

therefore,

$$f_2 = -\sum_{i=1}^{2} p_i \log_2 p_i - (1 - p_1 - p_2) \log_2 (1 - p_1 - p_2) \ . \tag{9}$$

Now let's observe its figure as shown in Fig.2.

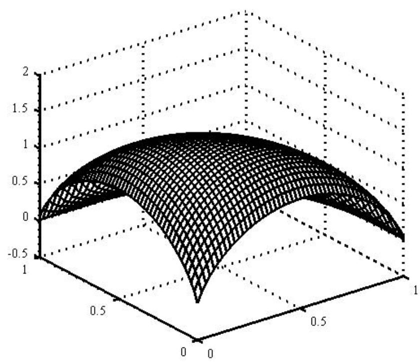

Fig. 2. Entropy function with two variables. It's a convex function with one maximum.

It is obvious that f_2 is also a convex function. We can find its maximum at point $(1/3, 1/3)$. If the domain of (p_1, p_2) doesn't include the point $(1/3, 1/3)$, as discussed above, it can be inferred similarly that the smaller the distance between (p_1, p_2) and $(1/3, 1/3)$ is, the bigger f_2 evaluates. Now we investigate p_3. Because $p_3 = 1 - p_1 - p_2$, for $\alpha \in [0, 1]$,

$$\begin{aligned} &\sum_{i=1}^{2} (p_i - 1/3)^2 + (p_3 - \alpha)^2 \\ &= \sum_{i=1}^{2} (p_i - 1/3)^2 + (1 - p_2 - p_1 - \alpha)^2 \ , \end{aligned} \tag{10}$$

when $\alpha = 1/3$, $f_2(p_1, p_2, p_3)$ is approaching its maximum as $\sum_{i=1}^{3} (p_i - 1/3)^2$ is approaching 0.

By analogy with above, using the convex feature of the functions and the existence of the extrema, we can conclude that for any $p_i \in [a_i, b_i] \subset [0, 1]$, $f(p_1, p_2, \cdots, p_n)$ reaches its maximum at the same point as $\sum_{i=1}^{n} (p_i - 1/n)^2$ reaches its minimum. Therefore, it can be inferred that the optimization problems (2) and (6) are equivalent.

This is the proof.

According to Shannon's information theory [10], an entropy function reaches its maximum in a equiprobable distribution. Entropy is a probabilistic measure of uncertainty. From some aspect, the equiprobable distribution $(1/n, 1/n, \cdots, 1/n)$ is the very source of the uncertainty. The closer a distribution is to the source, the bigger the value of entropy is. This is an interpretation to the theorem from the view of entropy.

3 The Mutative Scale Chaos Optimization Algorithm (MSCOA)

3.1 The Principle of Chaos Optimization

Chaos is a common phenomenon existing in nonlinear definite systems. Owing to its stochastic property, ergodicity and intrinsic regularity, global optimization methods based on chaos are widely applied in optimization problems, see [11,12,13,14,15].

The basic idea of the mutative scale chaos optimization algorithm is as follows. Firstly a sequence of chaotic variables is created by iterations. The sequence is used to check the whole solution space, which is called rough searching. Then according to the result of rough searching, a currently optimal solution is selected and the searching space is limited to a smaller one depending on the optimal solution, which is called precise searching. The MSCOA method incorporates advantages of both rough and precise searching, and achieves the goal of obtaining optimal solution quickly and effectively.

Chaos optimization is realized through chaotic variables created by chaotic mapping functions. For example, the Logistic function

$$y_{k+1} = uy_k(1 - y_k) \ , \tag{11}$$

where u is the chaotic parameter, $y_k \in (0, 1)$, $k = 0, 1, 2, \cdots$. When $u = 4$, the mapping is in totally chaotic state. Having been created, the chaotic variables need to be mapped into solution space as

$$x_k^i = a^i + (b^i - a^i) \cdot y_k^i \ , \tag{12}$$

where $i = 1, 2, \cdots n$, n is the number of function's variables. $[a^i, b^i]$ are domains of the variables. For a gray-scale image, the interval of threshold is usually $[0, 255]$.

3.2 Design of the Algorithm

The problem to be optimized is (6). According to the basic steps of the MSCOA method, we design an algorithm to search for the optimal threshold in a 2-d gray level space. The procedure is as follows:

Step 1: Initialization: $k = 0$, chaotic variables $y_0^i = y^i \in (0, 1)$, $r = 0$, $a_r^i = 0$, $b_r^i = 255$, where $i = 1, 2$, k is the chaos iteration counter and r is the iteration counter for shrinking spaces; optimal chaotic variables $(Y^1, Y^2) = (0, 0)$; initialize minimum F^* with a big value and current optimal threshold $(s^*, t^*) = (0, 0)$;
Step 2: Mapping chaotic variables (y_k^1, y_k^2) to threshold variables (s_k, t_k):

$$s_k = a_r^1 + y_k^1(b_r^1 - a_r^1), t_k = a_r^2 + y_k^2(b_r^2 - a_r^2) \ ;$$

Step 3: Calculate $F(s_k, t_k)$, if $F(s_k, t_k) < F^*$, then $(Y^1, Y^2) = (y_k^1, y_k^2)$, $(s^*, t^*) = (s_k, t_k)$, $F^* = F(s_k, t_k)$; else move on to next;
Step 4: Calculate $y_k^i = 4 \cdot y_k^i \cdot (1 - y_k^i)$, $i = 1, 2$, $k = k + 1$;
Step 5: Repeat step2-step4; if F^* remains unchanged in T_1 iterations, continue;
Step 6: Shrink the space for searching:

$$a_{r+1}^1 = s^* - \rho(b_r^1 - a_r^1), b_{r+1}^1 = s^* + \rho(b_r^1 - a_r^1),$$
$$a_{r+1}^2 = t^* - \rho(b_r^2 - a_r^2), b_{r+1}^2 = t^* + \rho(b_r^2 - a_r^2),$$

where $\rho \in (0, 0.5)$, and make sure the new space isn't beyond the original boundary:
if $a_{r+1}^i < a_r^i$, then $a_{r+1}^i = a_r^i$; if $b_{r+1}^i > b_r^i$, then $b_{r+1}^i = b_r^i$;
then make some modifications to (Y^1, Y^2),

$$Y^1 = (s^* - a_{r+1}^1)/(b_{r+1}^1 - a_{r+1}^1),$$
$$Y^2 = (t^* - a_{r+1}^2)/(b_{r+1}^2 - a_{r+1}^2).$$

Step 7: Let $y_k^i = Y^i$, repeat step2-step6; if F^* remains unchanged in T_2 iteration steps, output (s^*, t^*) and F^*.

This is the whole procedure of the algorithm, in which parametersρ, T_1, T_2 is to be adjusted to control accuracy and convergence rate for different applications.

4 Simulation Results and Analysis

To test the algorithm, we choose two 256×256 sized gray-scale images, the bacteria image and rice image. The programming tool is MATLAB v7.0, and the configuration of the computer is 1.5GHz host frequency with 512M memory.

The 2-D chaotic variables are mapped to $[l, L] \times [l, L]$, $0 \leq l < L \leq 255$, where L, l is the upper and lower boundary of gray levels. As to the selection of parameters in the algorithm, we choose $\rho = 0.4$, $T_1 = 300$, $T_2 = 3$. For comparison, the algorithm in Ref.[9] is also simulated. Segmentation results are shown in Fig.3 and Fig.4. From left to right are the original image, segmentation result by the method in Ref. [9], segmentation result by our method.

The different optimal thresholds, minimum of object functions and time cost are compared in Table.1. Due to the randomicity of chaotic algorithms, we execute the algorithm 50 times, and get the best thresholds, minimal function values and average time cost. Result I is computed by Ref. [9] and II by our paper.

Table 1. Segmentation results by two methods, compared by optimal thresholds, extrema of the entropy functions, and average time cost

		Thresholds	Minimum	Time Cost (s)
I	Rice	(126,137)	2.9075	39.87
	Bacteria	(112,116)	3.374	35.23
II	Rice	(115,123)	9.85e-3	0.98
	Bacteria	(94,106)	1.089e-2	1.21

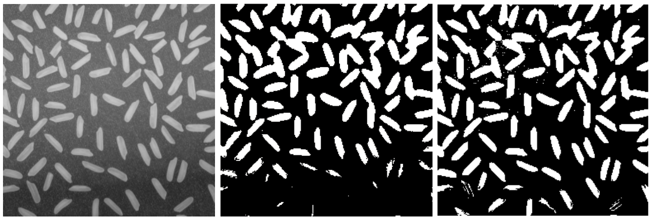

Fig. 3. Rice image and its segmentation results by two different methods. From left to right it's the original image, result by Ref.[9], result by our paper.

Fig. 4. Bacteria image and its segmentation results by two different methods. From left to right it's the original image, result by Ref.[9], result by our paper.

According to the segmentation results, it can be concluded that the algorithm by Ref.[9] is not likely to find the best threshold while our algorithm shows better segmentation results. It's not reasonable to judge the two algorithms by the minimums of two functions, for they are calculated in different methods. According to the time cost, algorithm in our paper shows a notable advantage over the other. As we know, the key point is how to search the 2-D threshold space. The algorithm in Ref.[9] is to search the space $(L-l)^2$ times, and $(L-l)^2$ calculations are executed for each feasible solution(s,t). So the computational complexity is $O[(L-l)^4]$. In our algorithm, according to parameters T_1 and T_2, only about 1000 iterations are executed and for every iteration $2(L-l)^2$ computations are needed. Thus the whole computational amount is $O[c\cdot(L-l)^2]$, which is much smaller than the former.

It is shown in Ref.[9] by experiment that the computational efficiency is increased by 15%-30% if function (3) is adopted comparing to function (1). However, out experimental results are not consistent with it. We test function (1) and (5) both in the MSCOA

Table 2. Two functions are both tested in the MSCOA method, compared by average iterations, the convergence rate, and average time cost

		Iterations	Convergence Rate	Time Cost (s)
I	Rice	1320	92%	4.88
	Bacteria	1171	90%	5.23
II	Rice	1294	94%	0.98
	Bacteria	1186	90%	1.21

method. The results are shown in Table.2 where I and II represents function(1) and (5) respectively.

It can be observed that the adoption of function (5) increases the computational efficiency about 4 times than function (1). The number of iterations is decided by MSCOA itself and has nothing to do with the adoption of functions. But the computational time changes noticeably, because for each iteration the time for calculating function (1) and (5) differs. Therefore, the application of the improved entropy function (5) is practical in 2-D entropic image segmentation.

By far people have been paying more attention to the genetic algorithm than chaos optimization algorithm. Theoretically both algorithms will converge to the global optimal solution, provided parameters are selected properly.Here the parameters are chosen according to Ref.[5]. There is a large body of literature focusing on the parameters for GA. Here the preferences may not be the best, but the result could illustrate something.The tool we use is Matlab v7.0, GA Toolbox.

Table 3. Comparison of GA and MSCOA in average iterations, convergence rate and time cost. I for GA and II for MSCOA. Notice that the iterations for GA is the number of generations.

		Iterations	Convergence Rate	Time Cost (s)
I	Rice	39	93.5%	2.24
	Bacteria	43	92%	2.67
II	Rice	1294	94%	0.98
	Bacteria	1186	90%	1.21

We know the GA is highly efficient for large scale optimization problems. Actually the problem here is a two dimensional case, seeming to be a little small for GA. If a segmentation method with three or more thresholds is used, GA may be a much better choice. Besides, the threshold is an integer in interval $[0, 255]$, which is very easy for binary coding and decoding. That's one of the reasons why people like to apply GA in image segmentation.

5 Conclusions

This paper presents an improved entropy function in 2-d entropic image segmentation. Compared to the original entropy function, the new one is simpler and easier for calculations. A mutative scale chaotic optimization algorithm is designed. Simulation results and comparison with other algorithms shows our algorithm is better. The MSCOA shows an excellent searching ability in our experiment, and it's easy for programming and fast in computation for such a 2-d entropic image segmentation.

Acknowledgement

The authors thank for the support of the National Natural Science Foundation of China, No.10571024.

References

1. Kapur, J.N., Sahoo, P.K., Wong, A.K.C.: A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision, Graphics, and Image Processing 29(3), 273–285 (1985)
2. Abutaleb, A.S.: Automatic thresholding of gray-level pictures using two-dimensional entropy. Computer Vision, Graphics, and Image Processing 47(1), 22–32 (1989)
3. Wang, L.S., Ou, Z.Y.: Image segmentation based on optimal histogram threshold by improved genetic algorithms. Journal of Data Acquisition and Processing 20(2), 130–134 (2005)
4. Wang, X., Wong, B.S., Tui, C.G.: X-ray image segmentation based on genetic algorithm and maximum fuzzy entropy. Robotics, Automation and Mechatronics, IEEE Proceedings 2, 991–995 (2004)
5. Lu, X.Q., Li, N., Chen, S.F., Ye, Y.K.: Two dimensional thresholding and genetic algorithms in image segmentation. Computer application and Software 18(12), 57–59 (2001)
6. Xiu, C.B., Liu, X.D., Zhang, Y.H.: Optimal entropy thresholding image segmentation based on chaos optimization. Computer Engineering and Application 27(2), 76–78 (2004)
7. Jansing, E.D., Albert, T.A., Chenoweth, D.L.: Two-dimensional entropic segmentation. Pattern Recognition Letters 20, 329–336 (1999)
8. Pal, N., Pal, S.K.: Object-background segmentation using new definitions of entropy. IEEE Proceedings 136(4), 284–295 (1989)
9. Yang, S., Gao, L.Q., Bian, L.Y.: Improvement of 2-d maximum entropy threshold algorithm based on optimal entropy function. Journal of System Simulation 17(6), 1350–1352 (2005)
10. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Tsinghua University Press, Beijing (2003)
11. Fujita, T., Watanabe, T., Yasuda, K., Yokoyama, R.: Global optimization method using chaos in dissipative system. Industrial Electronics, Control, and Instrumentation, IEEE Transactions 2(2), 817–822 (1996)
12. Zhang, H.M., Yang, J.M.: Improvement and application of mutative scale chaos optimization algorithm. Control and Decision 17(6), 598–601 (2002)
13. You, Y., Wang, S.A., Sheng, W.X.: New chaos optimization algorithm with applications. Journal of Xi'an Jiaotong University 37(1), 69–72 (2003)
14. Tokuda, I., Aihara, K., Nagashima, T.: Adaptive annealing for chaotic optimization. Physical Review E 58(4), 5157–5160 (1998)
15. Chen, L.N., Aihara, K.: Global searching ability of chaotic neural networks. Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions 46(8), 974–993 (1999)