

An Improved Image Segmentation Method based on Shannon Entropy and Biogeography based Optimization

Mengqing Feng^{1, 2, a}

¹School of Information Engineering, Zhengzhou University of Industrial Technology, Zhengzhou Henan 451150, PR China

²Machine learning and Data researching Institute, Zhengzhou University of Industrial Technology, Zhengzhou Henan 451150, PR China

^a903901419@qq.com

Keywords: Biogeography based optimization; Evolutionary algorithms; Multilevel thresholding; Image segmentation; Shannon entropy

Abstract. For the purpose of improve the effect of multilevel thresholding image segmentation, a new evolutionary optimization algorithm based on the science of biogeography for global optimization has been bring in namely Biogeography based optimization (BBO). In this paper we propose an improvement to BBO. In order to improve the diversity of population and to enhance its exploration ability, the Gaussian mutation operator is integrated into biogeography based optimization (BBO). And we combine this improved evolutionary algorithm and Shannon entropy to get multilevel thresholds of image segmentation. Experiments have been conducted on several images and compared with other algorithm namely ABC and DE. Simulation results and comparisons demonstrate the proposed BBO algorithm is better in terms of the quality of the solutions obtained.

Introduction

Image segmentation is one of the most important steps in the image processing [1-3]. It is a process that divides a raw input image into a number of non-overlapping regions such that each region is homogeneous and the union of two adjoining regions is heterogeneous and extracts interested region from the others. Numerous techniques have been developed for image segmentation to date. Thresholding is one of the most used techniques in image segmentation because of its computationally efficient. The objective of thresholding is to identify a threshold for (bi-level thresholding) or several thresholds (for multilevel thresholding) to subdivide the image into several levels. However, the computation gets more and more complex when multilevel thresholding extended [4]. Generally the exact locations of distinct valleys in a multimodal histogram of an image which can segment the image efficiently is not easy to identify. Thus to solve the problem of multilevel thresholding is a very important research throughout the world. At present, there are several algorithms have been proposed in literatures that have solved the issue of histogram thresholding. Kapur, Sahoo, and Wong proposed a method for gray-level image thresholding using the entropy of the histogram [5]. Horng (2010) used the honey bee mating optimization (HBMO) to search for the thresholds of histogram of image [6]. In this paper a thresholding method based on maximum entropy and the improved BBO [7-8] is presented, it uses maximum entropy criteria alone with the improved BBO to select the optimal threshold value. Due to the application of the preposed optimization, the computational cost are decreased and segmentation performance are enhanced.

The paper is organized as follows: In Section 2, Shannon entropy criterion based measure is described. Section 3 explains the improved BBO. Section 4 describes image segmentation based on the proposed method. In Section 5, experimental results on three images are presented. Conclusions are drawn in Section 6.

(Shannon) Entropy Criterion Based Measure

The Kapur method namely entropy criterion has been broadly used in estimating whether the characteristics provided by optimal thresholding method are satisfactory to the histogram-based image segmentation. The original algorithm in Kapur et al.(1985) has been developed for bi-level thresholding and can be described as follows.

Let there be L gray levels in a given image and these gray levels are in the range $\{0,1,2,\dots,(L-1)\}$. Then one can define $P_i = h(i)/N$, $(0 \leq i \leq (L-1))$ where $h(i)$ represents number of pixels with gray-level i , the total number of pixels in the image are represented by N , and $N = \sum_{i=0}^{L-1} h(i)$.

Bi-level Thresholding. The bi-level thresholding problem can be described as follows:

maximize $f(t) = H_1 + H_2$, where (1)

$$H_1 = - \sum_{i=0}^t \frac{P_i}{p_1(t)} \ln \left(\frac{P_i}{p_1(t)} \right) = \ln(p_1(t)) - v_1 / p_1(t), \quad p_1(t) = \sum_{i=0}^t P_i, \quad v_1 = \sum_{i=0}^t P_i \ln(P_i) \quad (2)$$

$$H_2 = - \sum_{i=t+1}^{L-1} \frac{P_i}{p_2(t)} \ln \left(\frac{P_i}{p_2(t)} \right) = \ln(p_2(t)) - v_2 / p_2(t), \quad p_2(t) = \sum_{i=t+1}^{L-1} P_i, \quad v_2 = \sum_{i=t+1}^{L-1} P_i \ln(P_i) \quad (3)$$

the optimal threshold is the gray level which maximizes $f(t)$.

Multilevel Thresholding. This entropy criterion based measure has later been extended to solve multilevel thresholding problems. It can be described as follows. Here, given a problem of determining c thresholds for a given image $[t_1 t_2 \dots t_c]$, the objective is to maximize the following function:

$$f([t_1 t_2 \dots t_c]) = H_1 + H_2 + \dots + H_c \quad (4)$$

where

$$\begin{aligned} p_1 &= \sum_{i=0}^{t_1} P_i, \quad H_1 = - \sum_{i=0}^{t_1} \frac{P_i}{p_1} \ln \left(\frac{P_i}{p_1} \right), \\ p_2 &= \sum_{i=t_1+1}^{t_2} P_i, \quad H_2 = - \sum_{i=t_1+1}^{t_2} \frac{P_i}{p_2} \ln \left(\frac{P_i}{p_2} \right), \\ p_3 &= \sum_{i=t_2+1}^{t_3} P_i, \quad H_3 = - \sum_{i=t_2+1}^{t_3} \frac{P_i}{p_3} \ln \left(\frac{P_i}{p_3} \right), \\ p_c &= \sum_{i=t_{c-1}+1}^{L-1} P_i, \quad H_c = - \sum_{i=t_{c-1}+1}^{L-1} \frac{P_i}{p_c} \ln \left(\frac{P_i}{p_c} \right) \end{aligned} \quad (5)$$

The Improved BBO

Biogeography-Based Optimization. Biogeography-based optimization is a new evolutionary optimization algorithm for global optimization which can be very useful in many applications [9-11]. The habitats in BBO algorithm are individual units, each habitat represents a solution to the problem. Features that used to describe the habitat include such factors as temperature, diversity of vegetation, land area, diversity of topographic features, and rainfall are called suitability index variables (SIVs), represent several components of the solution. Features represented by vectors.

Habitat suitability index (HSI) described the important characters of each habitat, represent the capacity of species of a habitat. SIVs can be considered as the independent variables of the habitat, and HSI can be considered as the dependent variable. HIS corresponds to the fitness function of the optimization problems. Migration and mutation are the most important measures in this optimization.

Habitats with a high HSI tend to have a better biogeographical environment and a large number of species, while those with a low HSI have a small number of species. Habitats with a high HIS represents good solutions ,while those with a low HIS analogous to poor solutions. according to the science of biogeography, habitats which suitable for living are more static in their species distribution. because of the large number of species they host, the high HIS habitats have many species that emigrate to nearby habitats while few species immigrate in. namely, high HIS habitats have a high emigrate rate and a low immigrate rate. In the same way, low HIS habitats have a low emigrate rate and a high immigrate rate. The immigrate and emigrate of species provided means of information exchange for habitats.,which is similar to the information exchange in different solutions in the optimization.

The habitats in BBO algorithm represent some candidate solutions to a problem. A good solution is analogous to an habitat with a high HSI, and a poor solution represents an habitat with a low HSI. High HSI solutions resist change more than low HSI solutions. In the same way, high HSI solutions tend to share their features with low HSI solutions. Thus,may improve the quality of poor solutions.

The mutation of BBO algorithm came from evolutionary computation. The mutation strategy was recommended in order to prevent optimization limited to local optimum. Changing one or more features of the habitat such that HSI varies from before.In the same way,mutating one or more components of the solution can update the value of HSI.

The BBO algorithm can be informally described with the following algorithm.

Initialize the BBO parameters. Initialize the number of habitats N , dimensions of the problem to be optimized D , and the maximum species count $S(\max)$ of habitats. We also initialize the maximum migration rates E and I , the maximum mutation rate m_{\max} , and elitism parameter.

Initialize a random set vector of habitats, each habitat corresponding to a potential solution to the given problem.

For each habitat, map the HSI to the number of species S , the immigration rate λ , and the emigration rate μ .

Probabilistically use immigration and emigration to modify each non-elite habitat.

For each habitat, update the probability of its species count. Then, mutate each non-elite habitat based on its probability, and recompute each HSI.

Go to step (3) for the next iteration. This loop can be terminated after a predefined number of generations or after an acceptable problem solution has been found.

The Improved BBO Algorithm. Migration and mutation are basic operators of BBO algorithm. We use migration to probabilistically share information between habitats. In this way, poor solutions can get some features from good solutions to improve its quality. Mutation can achieve greater change of the component within a certain range, enhance the search capabilities of the solution space. Similar to the mutations of biological evolution, mutation operator is important to promote the capability of optimization.

In order to enhance the exploration ability of BBO, we proposed a classical mutation strategy, Gauss mutation.The mutations varies the value from time to time generated Gaussian distribution. Two parameters that affect the Gaussian distribution are mean and variance in Gaussian distribution. If the range of the component to be mutation is symmetry,such as $[-X,X]$,then the mean value of 0,and variance value of 1. The variance determine the range of each mutation.

The Gaussian mutation can be described as follows:

The formula for the probability density function of the Guassian distribution is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Where μ is the mean and σ^2 is the variance. Then the Gaussian mutation with $\mu=0$ and $\sigma=1$ can be described as

$$x_{i,j} = x_{i,j} + N_j(0,1)$$

Where $x_{i,j}$ is j th dimension variable of individual X_i . And for each individual of j , the random number is generated a new one by $N_j(0,1)$.

Operational processes of Gauss mutation and random mutation are basically the same. Replace the random number of with Gaussian function as follows: Replace $x_{i,j}$ with a Gauss mutation to generate, $x_{i,j} = x_{i,j} + N_j(0,1)$.

The improved BBO algorithm can be informally described as the following algorithm.

Step1. Initialize the BBO parameters, including the habitat modification probability, mutation probability, maximum species count S , maximum migration rates E and I , maximum mutation rate m_{max} , and elitism parameter.

Step2. Initialize the generation counter: $g=0$.

Step3. Create a random initial population $X_i, x_{i,j} = \text{round}(l_j + (u_j - l_j)\text{rand}(0,1))$, $i=1, \dots, N$, $j=1, \dots, D$.

Step4. Evaluate $f(x_i)$.

Step5. **for** $g=1$ to g_{max} **do**

For $i=1$ to N **do**

Sort the population from the best fitness to the least one

Map the HIS to the number of species

Calculate the immigration rate λ_i and the emigration rate μ_i .

Modify the non-elite members of the population probabilistically with the migration operator.

End for

For $i=1$ to N **do**

Mutate the non-elite members of the population with the mutation operator.

End for

For $i=1$ to N **do**

Evaluate the new individuals in the population

Replace the habitats with their new versions

Replace the worst with the previous generation's elites

End for

$g = g+1$

End for

Step6. Migration: **for** $i=1$ to N **do**

For $j=1$ to D **do**

Use λ_i to probabilistically decide whether to immigrate to $y_{i,j}$.

If $\text{rand}(0,1) < \lambda_i$ **then**

Select the emigrating island X_k with probability $\propto \mu_k$

Replace the j th decision variable (SIV) of Y_i with its corresponding variable in X_j

$y_{i,j} = x_{k,j}$

End if

End for

End for

Step7. Mutation: **for** $i=1$ to N **do**

Compute the probability P_i using λ_i and μ_i

Use the probability P_i to compute the mutation rate m_i

For $j=1$ to D **do**

Select a variable (SIV) $x_{i,j}$ with a probability P_i

If $rand(0,1) < m_i$ **then**

Replace $x_{i,j}$ with a Gauss mutation to generate, $x_{i,j} = x_{i,j} + N_j(0,1)$

End if

End for

End for.

Step8. Go to Step4 for the next iteration. This loop can be terminated after a predefined number of generations or after an acceptable problem solution has been found.

Image Segmentation Based on the Proposed Method

Considering the BBO and maximum entropy method together into image segmentation, we set the threshold vector $X, x_i = [x_1, x_2, \dots, x_D]$, as the search variables, and formula (4) as the objective function f to guide the search. After getting the 1-D histogram of the image for each variable x_i , we adopt the improved BBO procedure to search the optimal result which can produce the maximum fitness. Then, the image can be segmented according to the values. On the basis of the above, the multilevel thresholding method based on maximum entropy and the improved BBO is realized as follows:

Step 1: Input an image and calculate its gray histogram, define a, b as the lower and upper boundaries of the threshold values. And calculate the probabilities of each grey level p_i with formula (5). Initialize the BBO parameters, including the habitat modification probability, mutation probability p_{mutate} , maximum species count S , maximum migration rates E and I , maximum mutation rate m_{max} , and elitism parameter K . Initialize the number of the habitats N , the maximum iterations $Maxgen$ and the number of the threshold D .

Step 2: Create a random initial population x_i , $x_{i,j} = round(l_j + (u_j - l_j)rand(0,1))$, $i=1, \dots, N$, $j=1, \dots, D$.

Step 3: Evaluate the HSI of each habitat with $f(x_i)$ and put the value into $x_{i,D+1}$. And according to the value of the HSI sort the X_i in descending order.

Step 4: for $g=1$ to $Maxgen$ **do**

For $i=1$ to N **do**

Sort the population from the best fitness to the least one, save the best habitats in a temporary array.

Map the HIS to the number of species. Calculate the immigration rate λ_i and the emigration rate μ_i .

Modify the non-elite members of the population probabilistically with the migration operator.

End for

For $i=1$ to N **do**

Mutate the non-elite members of the population with the mutation operator.

End for

For $i=1$ to N **do**

Evaluate the new individuals in the population

Replace the habitats with their new versions

Replace the worst with the previous generation's elites

End for

$g = g+1$

End for

Step 5: Migration: for $i=1$ to N **do**

For $j=1$ to D **do**

Use λ_i to probabilistically decide whether to immigrate to $y_{i,j}$.

If $\text{rand}(0, 1) < \lambda_i$ **then**

Select the emigrating island X_k with probability $\propto \mu_k$

Replace the j th decision variable (SIV) of Y_i with its corresponding variable in X_j

$y_{i,j} = x_{k,j}$

End if

End for

End for

Step 6: Mutation: for $i=1$ **to** N **do**

Compute the probability P_i using λ_i and μ_i

Use the probability P_i to compute the mutation rate m_i

For $j=1$ **to** D **do**

Select a variable (SIV) $x_{i,j}$ with a probability P_i

If $\text{rand}(0,1) < m_i$ **then**

Replace $x_{i,j}$ with a Gauss mutation to generate, $x_{i,j} = x_{i,j} + N_j(0,1)$

End if

End for

End for.

Step 7: Go to Step4 for the next iteration. This loop can be terminated after a predefined number of generations or after an acceptable problem solution has been found.

Step 8: Segment the image with the threshold values.

The flowchart of the procedure above is as flowchart1.

Experimental Results

We tested the proposed multilevel thresholding methods on several different standard test images. And we choose three of those pictures namely lena(512×512), house(256×256), pepper(512×512), and their test results to compare with DE, ABC algorithms. The optimized results of three optimization algorithms on three images in different thresholds are presented on Table 1. And picture 1, picture 2, picture 3 shown the original image, its histogram and segmented results of the three images. All the experiments are run on the Intel(R) AT/AT Compatible with CPU 1.8G and 256M DDR RAM and MATLAB 2012a.

To guarantee the accuracy of the test, the three algorithms were ran on the same condition and the same max_FES. From Table 1, it can be seen that the computation time of the proposed BBO algorithm is faster than DE, ABC algorithms. It used almost one third of the time that other algorithms used to got the same segmentation result. And the computation time increases along with the threshold level increases. Since the evolutionary methods are random and stochastic searching algorithms, it is necessary to analyze the stability of the evolutionary algorithms. According to the values of mean, std dev, max and min from Table 1, we can come to the conclusion that the proposed BBO method is as stable as DE and ABC methods. As indicate in Table 1, the proposed multilevel thresholding methods has the highest rate of success.

flowchart1

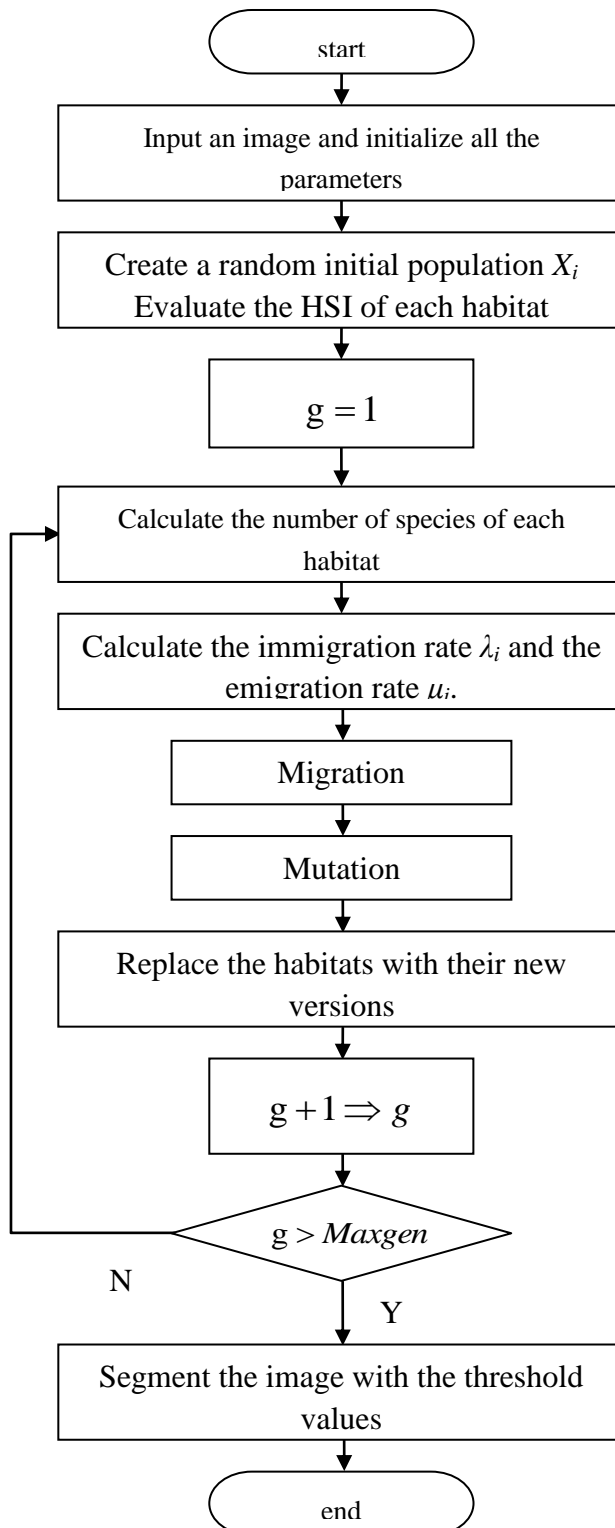


Table 1 optimized results of three optimization algorithms on three images in different thresholds

Image	thresholds	Max_FFS	Method	Time(s)	Mean	Std dev	Max	Min	SR
Lena.jpg (512×512)	2 (93,159)	6600	DE	0.5919	12.3883	0	12.3883	12.3883	100%
			ABC	0.6094	12.3883	0	12.3883	12.3883	100%
			BBO	0.2140	12.3883	0	12.3883	12.3883	100%
	3 (74,119,165)	7400	DE	0.6761	15.3612	1.8067e-15	15.3612	15.3612	100%
			ABC	0.7027	15.3612	1.8067e-15	15.3612	15.3612	100%
			BBO	0.2528	15.3612	1.8067e-15	15.3612	15.3612	100%
	4 (73,117,162,212)	8200	DE	0.7760	18.1397	3.3053e-4	18.1399	18.1386	91%
			ABC	0.7839	18.1399	8.5377e-5	18.1399	18.1395	98%
			BBO	0.2960	18.1393	3.4545e-3	18.1399	18.1210	99%
	5 (65,105,144,177,213)	9000	DE	0.8715	20.8308	4.3935e-3	20.8312	20.8298	88%
			ABC	0.8757	20.8310	3.6323e-4	20.8312	20.8297	95%
			BBO	0.3408	20.8312	0	20.8312	20.8312	100%
House.jpg (256×256)	2 (92,187)	6600	DE	0.5796	11.6079	1.8067e-15	11.6079	11.6079	100%
			ABC	0.5949	11.6079	1.8067e-15	11.6079	11.6079	100%
			BBO	0.2045	11.6079	1.8067e-15	11.6079	11.6079	100%
	3 (59,105,187)	7400	DE	0.6737	14.6239	0	14.6239	14.6239	100%
			ABC	0.6816	14.6239	0	14.6239	14.6239	100%
			BBO	0.2418	14.6239	0	14.6239	14.6239	100%
	4 (52,80,107,187)	8200	DE	0.7661	17.1126	1.0236e-3	17.1130	17.1087	91%
			ABC	0.7741	17.1129	1.8683e-4	17.1130	17.1120	96%
			BBO	0.2818	17.1109	7.8937e-3	17.1130	17.0819	98%
	5 (59,94,128,171,188)	9000	DE	0.8586	19.9332	1.1399e-3	19.9337	19.9287	93%
			ABC	0.8588	19.9333	1.0432e-3	19.9337	19.9301	93%
			BBO	0.3238	19.9337	7.2269e-15	19.9337	19.9337	100%
Pepper.jpg (512×512)	2 (60,135)	6600	DE	0.5891	12.4449	0	12.4449	12.4449	100%
			ABC	0.6032	12.4449	0	12.4449	12.4449	100%
			BBO	0.2122	12.4449	0	12.4449	12.4449	100%
	3 (58,120,178)	7400	DE	0.6814	15.6493	7.2269e-15	15.6493	15.6493	100%
			ABC	0.6902	15.6493	7.2269e-15	15.6493	15.6493	100%
			BBO	0.2526	15.6493	7.2269e-15	15.6493	15.6493	100%
	4 (49,90,134,178)	8200	DE	0.7774	18.5508	3.6134e-15	18.5508	18.5508	100%
			ABC	0.7837	18.5508	1.2203e-5	18.5508	18.5508	100%
			BBO	0.2962	18.5508	3.6134e-15	18.5508	18.5508	100%
	5 (49,89,133,177,211)	9000	DE	0.8715	21.3686	1.1286e-4	21.3686	21.3682	98%
			ABC	0.8769	21.3686	4.1163e-5	21.3686	21.3684	98%
			BBO	0.3415	21.3686	3.6134e-15	21.3686	21.3686	100%

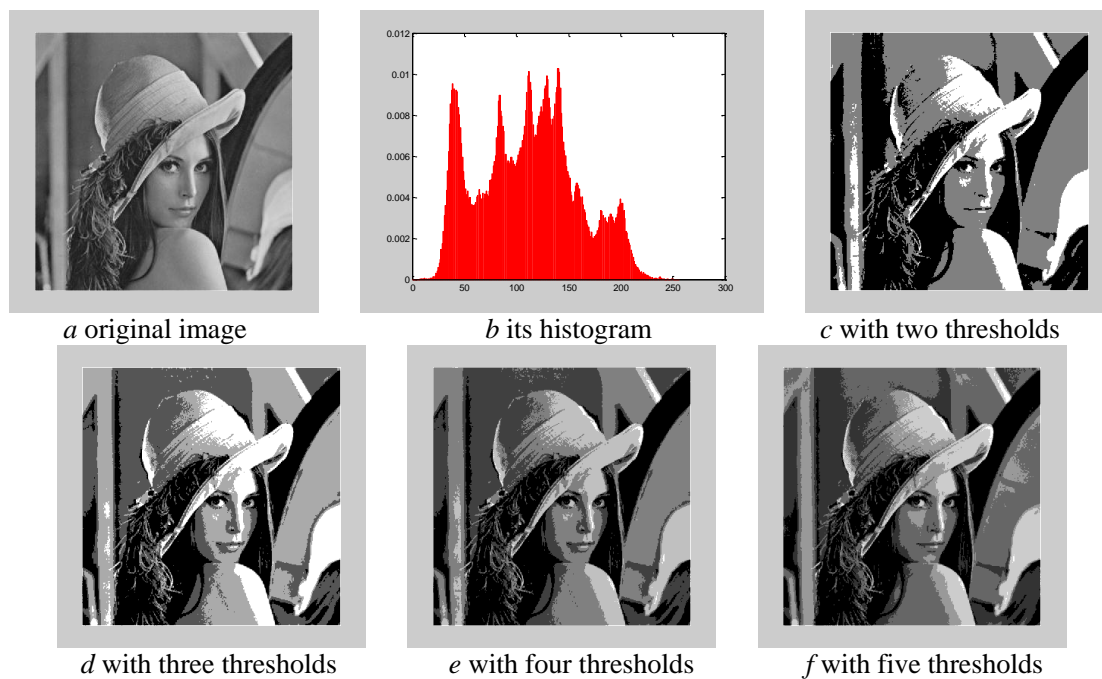


Figure 1. Lena image, its histogram and segmented results

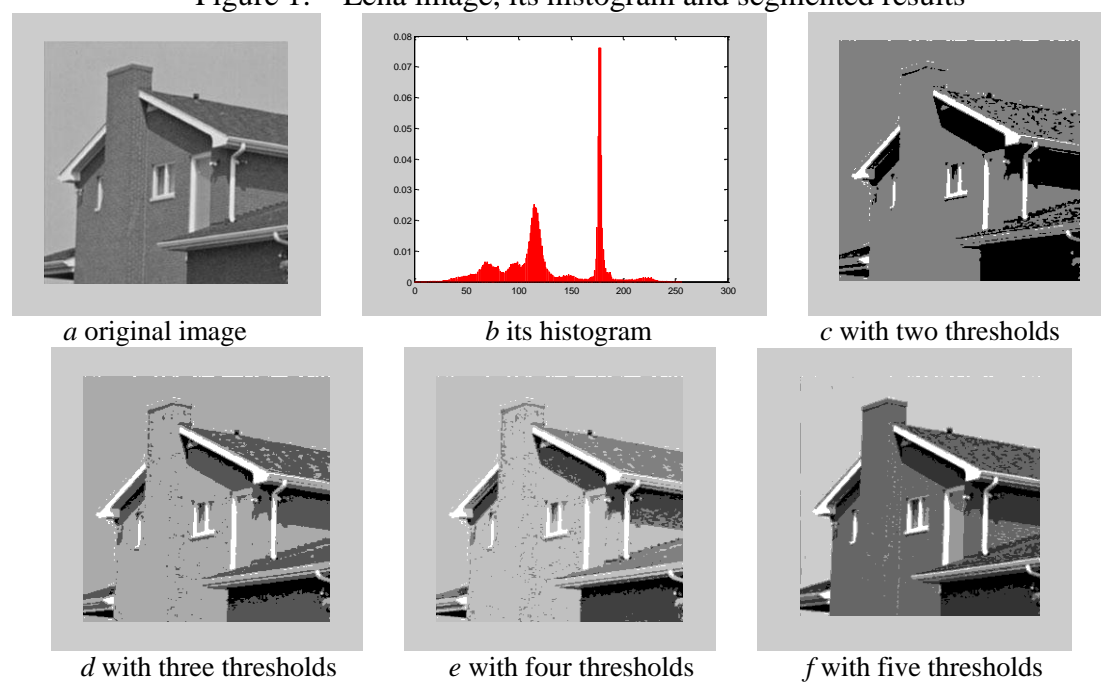
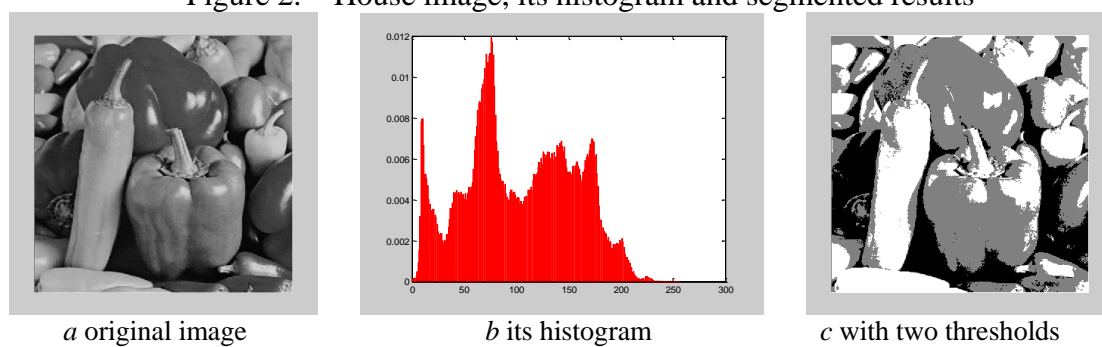


Figure 2. House image, its histogram and segmented results



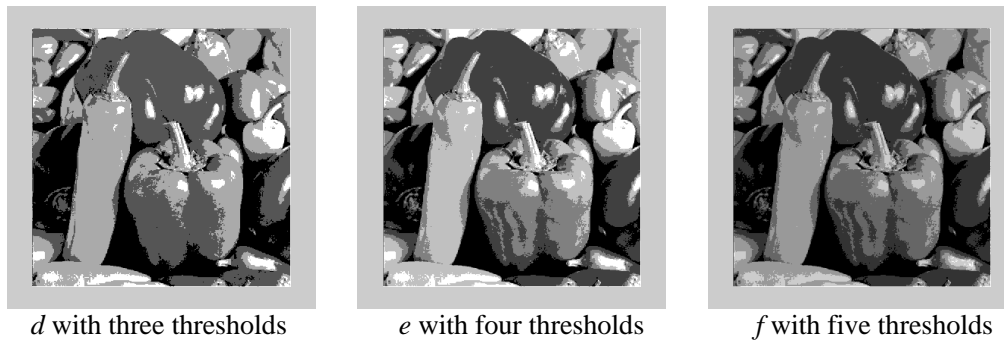


Figure 3. Pepper image, its histogram and segmented results

Conclusion

In this paper, we proposed a multilevel thresholding method based on Shannon entropy using BBO algorithm for image segmentation. And we recommend Gaussian mutation into the mutation process of the BBO algorithm to enhance the exploration ability. The performance of the proposed algorithm has been test on several different standard test images and compared with DE,ABC methods. The results of image segmentation show that the improved BBO algorithm is better than the other algorithms in terms of computation efficiency , solution quality and stability. More research is need to be done to test the practicability of the proposed algorithm on various types of images.

References

- [1] A.K. Bhandari, V.K. Singh , A. Kumar , et al. Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur's entropy[J]. Expert Systems with Applications, 2014, 41(7): 3538-3560.
- [2] P. Li, Z. Li. Color image segmentation using PSO-based histogram thresholding [J]. WIT Transactions on Information and Communication Technologies, 2014, 52: 1601-1607.
- [3] M. Horng. Multilevel thresholding selection based on the artificial bee colony algorithm for image segmentation [J]. Expert Systems with Applications, 2011, 38(11):13785-13791.
- [4] P.D. Sathya, R. Kayalvizhi, Modified bacterial foraging algorithm based multilevel thresholding for image segmentation. Engineering Applications of Artificial Intelligence, 24 (2011) ,pp.595–615.
- [5] J. N. Kapur, P. K. Sahoo and A. K. C. Wong. A new method for gray-level picture thresholding using the entropy of the histogram. Computer Vision Graphics Image Processing, 1985, 29, pp. 273-285.
- [6] M. H. Horng. A multilevel image thresholding using the honey bee mating optimization. Applied Mathematics and Computation, 215, 3302–3310.
- [7] Dan Simon, Senior Member, Biogeography-Based Optimization. IEEE Transactions on Evolutionary computation,2008,12(6),pp.702–713.
- [8] W. Y. Gong, Z. H. Cai, C. X. Ling, H. Li. A real-coded biogeography-based optimization with mutation [J].Applied Mathematics and Computation, 2010, 216(9): 2749-2758.
- [9] A. Chatterjee, P. Siarry, A .Nakib, et al. An improved biogeography based optimization approach for segmentation of human head CT-scan images employing fuzzy entropy[J]. Engineering Applications of Artificial Intelligence, 2012, 25(8): 1698-1709.
- [10] Rahmati S H A, Zandieh M. A new biogeography-based optimization (BBO) algorithm for the flexible job shop scheduling problem [J]. The International Journal of Advanced Manufacturing Technology, 2012, 58(9-12): 1115-1129.

- [11] A. Bhattacharya, P. K. Chattopadhyay. Hybrid differential evolution with biogeography-based optimization algorithm for solution of economic emission load dispatch problems [J]. Expert Systems with Applications, 2011, 38(11): 14001-14010.