Yongquan Zhou[1,2], Zhe Ouyang[1] , Jiakun Liu[1], Gaoli Sang[1]

Guangxi University for Nationalities (1), Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis (2)

A Novel K-means Image Clustering Algorithm Based on Glowworm Swarm Optimization

***Abstract**. In this paper, the classical K-means clustering algorithm disadvantages are given, a novel image clustering algorithm based on glowworm swarm optimization (ICGSO) is proposed. The new image clustering algorithm has been applied to several benchmark images to illustrate its applicability. The experimental results show that the proposed algorithm not only avoids the local optima, but also obtains the better image classification effects.*

***Streszczenie.**.W artykule przedstawiono niedogodności klasycznej metody otrzymywania klastrów przy przetwarzaniu obrazów i zaproponowano nową metodą bazującą na optymalizacji z wykorzystaniem algorytmów świetlikowych. (**Nowa metoda klastrowania obrazów bazująca na optymalizacji z wykorzystaniem algorytmów świetlikowych**)*

Keywords:. glowworm swarm optimization; k-means clustering algorithm; image clustering
Słowa kluczowe: przetwarzanie obrażów, klastry, algorytmy świetlikowe.

Introduction

Image classification is an image processing method of distinguishing between different categories of objectives, according to the different features contained in the image information. It is pattern recognition's application in the field of image processing. In the same conditions, similar objects of image should get the similar spectral information and spatial information, and show some inherent similarity of the similar objects. It means the feature vectors of the similar objects pixels will cluster in the spatial areas with the same features, but the feature vectors of the different objects pixels will cluster in the spatial areas with different features [1-3]. The image classification and clustering methods can be mainly divided into two kinds, supervised clustering and unsupervised clustering. In the supervised approach, the number and the numerical characteristics of the classes in the image are known in advance (by the analyst) and used in the training step, which is followed by a classification step. Unsupervised classification is in no prior knowledge of the category, through the measure of pixels can be classified as the attribute different types of classification method, and unsupervised classification is also referred to as a clustering problem [13].

K-means algorithm [4] proposed by MacQueen is a classical algorithm for clustering. It works through several iterations, and updates every cluster center gradually until getting the best clustering results. However, there are two drawbacks for this algorithm. It depends on the initial condition, which may cause the algorithm to converge to suboptimal solutions; and it falls into local optimum easily. To overcome the above shortcomings, many methods have been proposed to improve this algorithm in recent years, which gets certain achievements. But there is still phenomenon of premature convergence in the local optimum [5, 6].

Glowworm Swarm Optimization (GSO) [7-9] is a new method of swarm intelligence raised by K.N.Krishnanad and D.Ghose in 2005. This algorithm was inspired the phenomenon that the glow attracts mates. And the brighter the glow, more is the attraction. Each agent in the swarm decides its direction of movement by the strength of the signal picked up from its neighbors. Therefore, we use the glowworm metaphor to represent the underlying principles of our optimization approach. The GSO can do well in global searching, searching for the optimal clustering in parallel. Therefore, it can avoid the influence of the initial condition. This paper presents a new image clustering algorithm based on glowworm swarm optimization (ICGSO), and applies them to image classification. In this way, the basic glowworm swarm optimization was combined with the K-means algorithm. The new algorithm could overcome the disadvantages of K-means algorithm effectively and get better clustering qualities.

The rest of this paper is arranged as follows. In section 2, we discuss the basic GSO algorithm. The new image clustering algorithm based on glowworm swarm optimization (ICGSO) was described in section 3. And section 4 gives the experimental results. At last a conclusion in drawn in section 5.

Glowworm Swarm Optimization Algorithm (GSO)

In GSO [7-9], each glowworm distributes in the objective function definition space. These glowworms carry own luciferin respectively, and has the respective field of vision scope called local-decision range. Their brightness concerns with in the position of objective function value. The brighter the glow, the better is the position, namely has the good target value. The glow seeks for the neighbor set in the local-decision range, in the set, a brighter glow has a higher attraction to attract this glow toward this traverse, and the flight direction each time different will change along with the choice neighbor. Moreover, the local-decision range size will be influenced by the neighbor quantity, when the neighbor density will be low, glow's policy-making radius will enlarge favors seeks for more neighbors, otherwise, the policy-making radius reduces. Finally, the majority of glowworm return gathers at the multiple optima of the given objective function.

Each glowworm i encodes the object function value $J(x_i(t))$ at its current location $x_i(t)$ into a luciferin value l_i and broadcasts the same within its neighborhood. The set of neighbors $N_i(t)$ of glowworm i consists of those glowworms that have a relatively higher luciferin value and that are located within a dynamic decision domain and updating by formula (1) at each iteration.

Local-decision range update:

$$r_d^i(t+1) = \min\{r_s, \max\{0, r_d^i(t) + \beta(n_t - |N_i(t)|)\}\}; \tag{1}$$

and $r_d^i(t+1)$ is the glowworm i's local-decision range at the $t+1$ iteration, r_s is the sensor range, n_t is the neighborhood threshold, the parameter β affects the rate of change of the neighborhood range.

The number of glows $N_i(t)$ in local-decision range:

(2) $$N_i(t) = \{ j : \|x_j(t) - x_i(t)\| < r_d^{\,i}; l_i(t) < l_j(t)\};$$

and, $x_i(t)$ is the glowworm i 's position at the t iteration; $l_i(t)$ is the glowworm i 's luciferin at the t iteration; the set of neighbors of glowworm i consists of those glowworms that have a relatively higher luciferin value and that are located within a dynamic decision domain whose range r_d^i is bounded above by a circular sensor range $r_s(0 < r_d^i < r_s)$ .Each glowworm i selects a neighbor j with a probability $p_{ij}(t)$ and moves toward it. These movements that are based only on local information, enable the glowworms to partition into disjoint subgroups, exhibit a simultaneous taxis-behavior toward and eventually co-locate at the multiple optima of the given objective function.

Probability distribution used to select a neighbor:

(3) $$p_{ij}(t) = \frac{l_j(t) - l_i(t)}{\sum_{k\in N_i(t)} l_k(t) - l_i(t)};$$

Movement update:

(4) $$x_i(t+1) = x_i(t) + s\left(\frac{x_j(t) - x_i(t)}{\|x_j(t) - x_i(t)\|}\right);$$

Luciferin-update:

(5) $$l_i(t) = (1-\rho)l_i(t-1) + \gamma J(x_i(t));$$

and $l_i(t)$ is a luciferin value of glowworm i at the t iteration, $\rho \in (0,1)$ leads to the reflection of the cumulative goodness of the path followed by the glowworms in their current luciferin values; the parameter γ only scales the function fitness values, $J(x_i(t))$ is the value of fitness function.

In general, the GSO algorithm is described as four parts: initial distribution, luciferin update, the individual's movement and local-decision range update.

K-means Image Clustering Algorithm Based on GSO

K-means Image clustering algorithm based on glowworm swarm optimization (ICGSO) was applied in image classification. A glowworm x_i can be expressed as a N_c -dimension vector $x_i = (z_{i1}, ..., z_{ij}, ..., z_{iNc})$, in which z_{ij} means the j -th cluster center vector of the i -th agent, and N_c is the number of clusters. So a swarm represents a set of candidates of cluster centers. Three ways can be used to measure the quality of image clustering algorithms: quantization error, the maximum intra-distance, the minimum inter-distance [10-13].

Quantization Error

Quantization error means the average Euclidean distance of all the pixels to their associated cluster centers, defined as

(6) $$Je = \frac{\sum_{j=1}^{Nc}\left[\sum_{\forall m_p \in c_{ij}} d(m_p, z_{ij})\right] / |C_{ij}|}{N_c}$$

where

(7) $$d(m_p, z_{ij}) = \sqrt{\sum_{l=1}^{N_b}(m_{pl} - z_{ijl})^2}$$

where N_c is the number of clusters, m_p denotes the N_b components of pixel p ,denotes gray value for grayscale images, c_{ij} denotes the j th cluster of the i -th agent, z_{ij} denotes the j th cluster center vector of the i -th glowworm, $|C_{ij}|$ is the cardinality of the set C_{ij} , $d(m_p, z_{ij})$ means the Euclidean distance of pixels to their associated classes, N_b denotes the number of spectral bands of the image set (1 for grayscale images).

The Maximum Intra-distance

The maximum intra-distance means the maximum average Euclidean distance of pixels to the associated cluster centers, using

(8) $$d_{\max}(x_i, M) = \max_{j=1,...,Nc}\left\{\sum_{\forall m_p \in C_{ij}} d(z_{ij}, m_p) / |C_{ij}|\right\}$$

where M is the glowworms clustering domain individual said the characteristic value of each pixel consists of the matrix.

The Minimum Inter-distance

The minimum inter-distance means the minimum Euclidean distance between any pair of clusters, defined as

(9) $$d_{\min}(x_i) = \min_{\forall j_1, j_2, j_1 \neq j_2}\left\{d(z_{ij1}, z_{ij2})\right\}$$

To get the better clustering qualities, the smaller values are needed for quantization error and the maximum intra-distance; respectively, and a larger value for the minimum inter-distance. Thus, the fitness function of clustering qualities should constructed to minimize Je and $d_{\max}(x_i, M)$ but maximize $d_{\min}(x_i)$, defined fitness function as

(10) $$f(x_i, M) = w_1 d_{\max}(x_i, M) + w_2(m_{\max} - d_{\min}(x_i))$$

where $m_{\max}$ is the maximum pixel value in the image set. w_1 and w_2 are defined by the user, the fitness function prefers long distances of clusters from original pixels and small distances between clusters. Usually, let $w_1 = w_2$ =0.5.

The framework of the ICGSO algorithm is given as below:

Step 1. Initialize parameters of $\rho, \gamma, \beta, s, l_0, m, n$;

Step 2. Initialize each glowworm to contain N_c randomly selected cluster centers;

Step 3. For each glowworm x_i

3.1). For each pixel m_p

Calculate $d\left(m_p, z_{ij}\right)$ for all C_{ij} using equations (7);

Assign m_p to C_{ij} where

$$d(m_p, z_{ij}) = \min_{\forall c=1,\ldots,N_c} \{d(m_p, z_{ic})\}$$

3.2). Calculate the fitness $f\left(x_i, M\right)$ using equation (10);

Step 4. For $iter = 1$ to $iter_{\max}$,

For each glowworm x_i

Updating luciferin value according to equation (5);
Selects conforms to the condition glowworm according to equation (2);
Using (3) to select the distribution $j\left(j \in N_i(t)\right)$, and updating with equation (4);
The K-means algorithm is applied to GSO of each iterations.

4.1). For each pixel m_p

Calculate $d\left(m_p, z_{ij}\right)$ for all C_{ij} using equations (7); Assign m_p to C_{ij} where

$$d(m_p, z_{ij}) = \min_{\forall c=1,\ldots,N_c} \{d(m_p, z_{ic})\}$$

4.2). Update the new cluster means;

Calculate the fitness $f\left(x_i, M\right)$ using equation (10);

Revision search radius by equation (1);

Step 5. Segment the image using the optimal clusters centers given by the best global glowworm.

4. Experimental Results

4.1. Experimental Environment

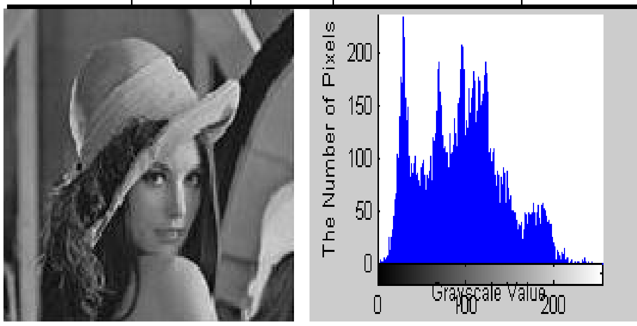
The K-means, fuzzy C-means method (FCM) and ICGSO are code in MATLAB7.0 and implemented on Intel Core2 T5300 1.73GHz machine with 1G RAM under windows XP platform. The set of ICGSO's parameters are as below $n = 10$, max of iteration $\max iter = 100$, $\rho = 0.4$, $\gamma = 0.6$, $\beta = 0.08$, moving step $s = 0.3$, $n_t = 5$ and initialization of luciferin $l_0 = 5$.

4.2 .The Test Images

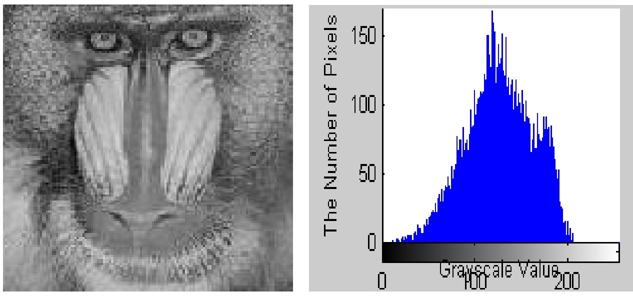
This section compares the results of the K-means algorithm, FCM algorithm and the new proposed algorithm on several benchmark images. They are Lena (134×140 8-bit), Mandrill (111×111 8-bit), Peppers (131×131 8-bit), MRI (108×120 8-bit).The color images are transferred to the gray scale images at first. The results have been stated in terms of the mean values and standard deviations over 20 independent runs in each case. Use Je, $d_{\max}(x_i, M)$ and $d_{\min}(x_i)$ to measure the clustering results, the results are shown in Table 1. K is the number of the clusters.

Table 1. Experimental of comparison

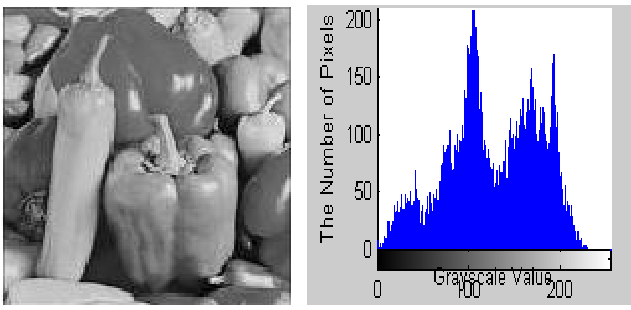
Image	Algorithm	Je	$d_{\max}(x_i, M)$	$d_{\min}(x_i)$
Lena ($K = 4$)	K-means	12.31	14.97	37.56
	FCM	11.87	17.82	34.36
	ICGSO	11.45	13.27	40.62
Mandrill ($K = 5$)	K-means	7.91	12.59	23.26
	FCM	7.62	12.81	21.73
	ICGSO	8.06	10.76	26.96
Peppers ($K = 5$)	K-means	9.92	14.59	28.39
	FCM	9.91	13.81	23.80
	ICGSO	9.28	10.50	32.29
MRI ($K = 6$)	K-means	10.11	15.78	34.03
	FCM	11.22	17.74	27.91
	ICGSO	9.41	12.63	32.68



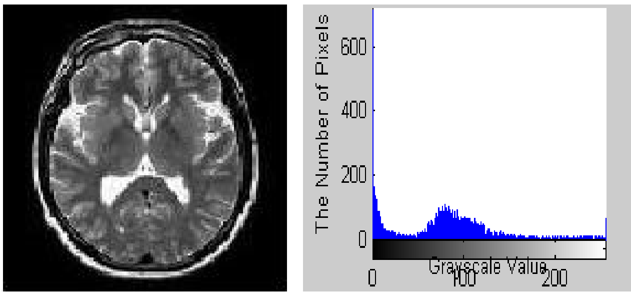
（a） Lena



（b） Mandrill



（c） Peppers

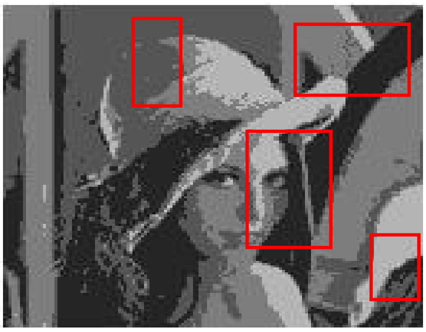


（d）MRI

Fig.1. Test images and gray scale histograms

K-means

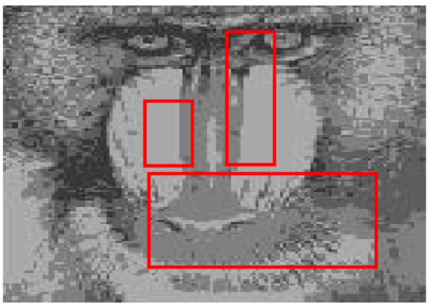
FCM



ICGSO

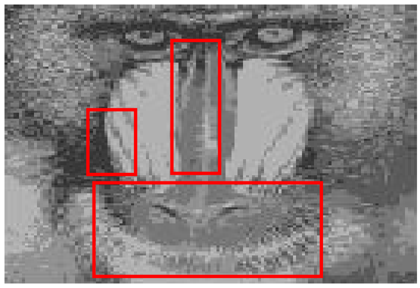
(a) Lena

（134×140；$K = 4$；$\max iter = 100$）



K-means

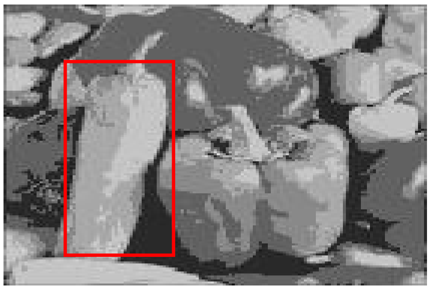
FCM



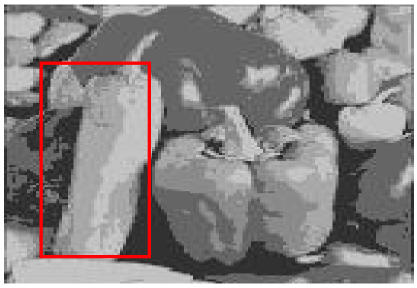
ICGSO

(b) Mandrill

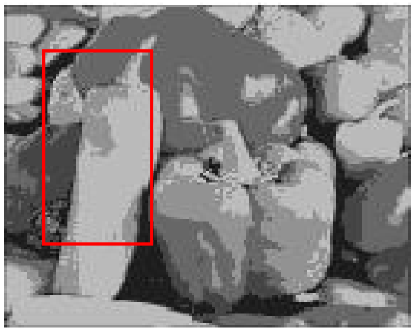
（111×111；$K = 5$；$\max iter = 100$）



K-means



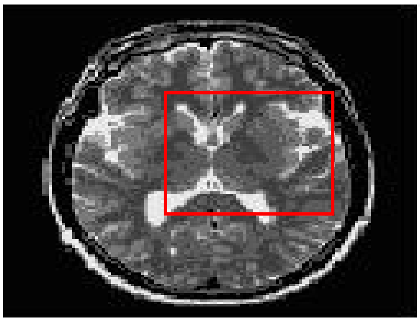
FCM



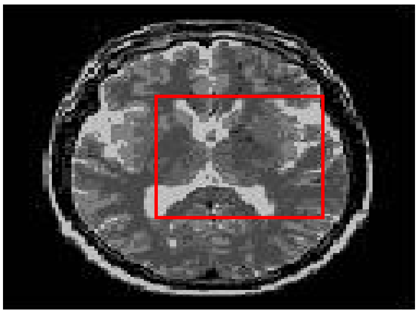
ICGSO

(c) Peppers

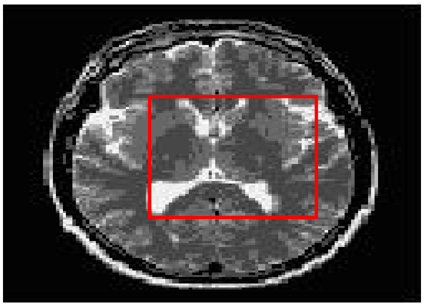
（131×131；$K = 5$；$\max iter = 100$）



K-means



FCM



ICGSO

（d）MRI

（108×120； $K=6$； $\max iter=100$）

Fig.2. The clustering results of K-means、FCM、ICGSO

Analyses of Results

Form Table 1 we can see that the proposed algorithm ICGSO outperforms the former two algorithms. For Lena and Peppers, the ICGSO can get the better values of quantization error, the maximum intra-distance and the minimum inter-distance than K-means and FCM. For Mandrill, the quantization error of ICGSO is a little larger than the other two algorithms, but the maximum intra-distance and the minimum inter-distance are better. For MRI, the minimum inter-distance of K-means algorithm is better. However, the performance of ICGSO is superior to FCM and K-means. Fig.2. shows the clustering results of three algorithms on four benchmark images respectively. Form the Fig.2. we can find that the clustering effect of ICGSO is better the former two on the four images. In general, ICGSO can cluster the images well.

Conclusions

In the paper, combining the GSO with K-means algorithm, we have presented K-means image clustering algorithm based on glowworm swarm optimization (ICGSO). The experimental results show this new algorithm performed very well when compared to the K-means algorithm. It's effective and more robust than the K-means algorithm and FCM algorithm. The researches and applications on GSO are still limited. We need to improve the searching speed and accuracy about the algorithm, and selecting the optimal clustering number is a question. That the GSO can be used widely in pattern recognition and image processing in the future.

Acknowledgements

This work is supported by National Science Foundation of China (61165015), Key Project of Guangxi Science Foundation (2012GXNSFDA053028), Key Project of Guangxi High School Science Foundation (20121ZD008) and the funded by open research fund program of key lab of intelligent perception and image under-standing of ministry of education of china under Grant (IPIU01201100).

REFERENCES

[1] Bicheng Zhou, Tianqiang Peng, Bo Peng. Intelligent image processing technology. Electronics Industy Press, (2004),285-319
[2] M.Sonka, V. Hlavac, R. Boyle. Image processing, analysis, and machine, (Second Edition). People Post Press, Beijing. (2003)
[3] Theodoridis S, Koutr oumbas K. Pattern Recognition. Berlin-Heidelber: Springer-Verlag, (2005),73-85
[4] Tou J T , Gonzalez R C. Pattern recognition principle. Addison wesley, Reading, (1974)
[5] Maulik U,B, Yopadhay S. Genetic algorithm-based cluster technique. Pattern Recognition. 33, (2000), 1455-1465
[6] Jigui Sun, Jie Liu, Lianyu Zhao.: Clustering algorithm research. Journal of Software. 19. (2008),No.9,125-133
[7] K.N.Krishnand, D.Ghose. Glowworm swarm optimization for simultaneous capture of multiple local optima of multimodal functions. Swarm Intell. 3, (2009),87-124
[8] K.N.Krishnand, D.Ghose. Glowworm Swarm Optimization: A new method for optimizing multi-modal functions. Int. J. Computational Intellingence Studies.1, (2009). 93-119
[9] Guotai Zeng. A glowworm algorithm for solving data clustering problems. in Department of information management, Tatung University, Taiwan.(2008),1-73
[10] Jingming Liu, Lichuan Han, Liwen Hou. Cluster analysis based on particle optimization algorithm. Systems Engineering Theory and Practice .6. (2005).,54-58.
[11] Shoubao Su, Jie Fang, Jiwen Wang. Image clustering methods based on invasive weeds optimization. Journal of South China University of Technology , 36, (2008), 95-100.
[12] Xiaogen Wang. Improved particle swarm optimization algorithm and its application in image. Jiangnan University, China. (2009),35-39
[13] Omran, M. G., Engelbrecht, A. P. and Salman, A particle swarm optimization method for image clustering. International Journal on Pattern Recognition and Artificial Intelligence. 19. (2005), 297-321

**Corresponding author: Yongquan Zhou, Ph.D.& Prof.*
College of Information Science and Engineering, Guangxi University for Nationalities, Nanning Guangxi, 530006, China
E-mail:yongquanzhou@126.com