Image Segmentation of Pseudo-foreign Fibers in Cotton on the Basis of Improved Genetic Algorithm

Lulu Ge, Daoliang Li, Liu Yang, Wenzhu Yang

Image Segmentation of Pseudo-Foreign Fibers in Cotton on the basis of Improved Genetic Algorithm

Lulu Ge[1] Daoliang Li[2,*] Liu Yang[1] Wenzhu Yang[3]

[1)] College of Engineering, China Agricultural University, Beijing, China (gelulugelulu@126.com)

[2)] College of Information and Electronic Engineering, China Agricultural University, Beijing, China (li.daoliang@163.com)

[1)] College of Engineering, China Agricultural University, Beijing, China (yangliu@cau.edu.cn)

[3)] College of Mathematics and Computer Science, Hebei University, Baoding, China (wenzhuyang@163.com)

Abstract. In the foreign fibers cleaning process, pseudo-foreign fibers are often mistaken for foreign fibers, this result not only seriously affects the detecting precision of foreign fibers cleaning machine, but also doubles the time of cleaning up lint. As for false identification problem of pseudo-foreign fibers in cotton, this paper proposes a new approach for fast segmentation of pseudo-foreign fibers in cotton on the basis of improved genetic algorithm. This improved genetic algorithm reduced the searching range for calculating optimal threshold from 0~255 to 100~220. The calculating speed in this stage was improved more than twice in average. The fitness amendments formula is also proposed to improve genetic algorithm disadvantage, at the same time, this solved issues of "premature", and converging to global optimal solution difficultly in the traditional algorithm. The results show that the algorithm has high speed, accuracy, anti-interference and so on.

Keywords: Cotton, Pseudo-foreign fibers, Genetic algorithm, Image segmentation,Threshold

1 Introduction

Sorting out foreign fibers in cotton is an urgent solved problem [1]. Foreign fibers have affected the quality of cotton products seriously [2]. Automated visual inspection (AVI) system is a main tool at present for real time foreign fibers detection in lint. The theory of this system is color differences of image processing technology [3-4]. The concept of color difference is based on that impurities of lint cotton and the cotton background have different colors. Hence, the impurities which has similar colors and shapes with foreign fibers were thought as foreign fibers to be removed, which is called as "pseudo-foreign fibers in cotton", also referred to be "pseudo-foreign fibers".

Pseudo-foreign fibers in cotton have various kinds, such as cottonseed, grass blade, boll shell, cotton leaf, cotton stick, weed, stringy cotton, stringy cotton, colored cotton and oiled cotton, etc. [5]. These pseudo-foreign fibers are very similar to the foreign fibers in color, size and shape. However, compared with foreign fibers, pseudo-foreign fibers harm hardly to the cotton into yarn, dyeing, bleaching and other aspects. On the new standard GB1103-2007 of cotton, these pseudo-foreign fibers in cotton are ruled to follow the normal process of impurities to remove, and then determine level according to standards cotton classification [6].

Meanwhile, the traditional methods of image segmentation, such as Otsu method, and the watershed method, are tested to process the pseudo-foreign fibers' image. The histogram information generated by these methods has only one peak, and the peak threshold is more concentrated, and segmentation effect is not satisfactory [7].

Due to false identification by cleaning machine of foreign fibers, pseudo-foreign fibers affected seriously the detection accuracy of the foreign fibers cleaning machine, also affected the efficiency of the foreign fibers cleaning machine. The time of cleaning up lint cotton is thousand times to increase, and also dragging processing progress of lint cotton.

At present, there is not pseudo-foreign fibers’ testing equipment in the world. Meanwhile, the pseudo-foreign fibers in cotton make cotton production and processing more difficultly. Hence, China Agricultural University and China Cotton Machinery & Equipment Co., Ltd. are jointly developing detection equipment on line of foreign fibers. Thereinto, Image segmentation is the key of improving the equipment measurement accuracy and reducing false recognition rate, so a suitable image segmentation algorithm of pseudo-foreign fibers determines the latter target recognition and classification of good or bad.

In order to overcome the shortcomings of the traditional methods and segment pseudo-foreign fibers target quickly, this paper puts forward to a new image segmentation method on the basis of improved genetic algorithm.

2 Samples and Methods

2.1 Sample Preparation

Due to the wide of pseudo-foreign fibers is very large, totally 150 kinds, so the typical pseudo-foreign fibers were chosen to do this research which included grass blade, cottonseed, stringy cotton, boll shell, cotton leaf, cotton stick, weed, stringy cotton, ginned dead cotton, waste paper, colored cotton and oiled cotton. Adequate pure cotton lint with no pseudo-foreign fibers was also prepared for making the lint layer. All these pseudo-foreign fibers' samples were provided by China Cotton Machinery & Equipment Co., Ltd.

2.2 Image Acquisition

Colorful images are captured by the color line scan camera of DALSA under high-brightness LED lighting. Shooting frame rate is 30fps @ 4000×500 color images. Image processing used the Pentium-M processor (CPU), clocked at 2.0GHz, 2GB of memory for the IPC, and software development tools for Visual C++ 6.0.

The steps of image acquisition are as follows. An opening machine is performed to make pseudo-foreign fibers away from the cotton layer as far as possible which is 80 cm wide and 2 mm thick, and then the images are acquired by the DALSA line scan camera. The capture card saves the images following the order of cotton layer passing by.

2.3 Method Selection

There are many image segmentation methods. According to the different treatment strategies, segmentation algorithms can be divided into serial and parallel algorithm. The serial algorithm determines the current point as edge depending on edge detection to get the result of the previous point, but the parallel algorithm only depends on the current point and its neighborhood points. Therefore, the edge detection operator can act on each pixel of the image in parallel algorithm, but the serial algorithm depends on start point and the previous point. That means parallel algorithm's judgments and decisions can be made independently and simultaneously. In the serial algorithm, the results of early treatment can be used by subsequent processing. Therefore, serial segmentation algorithm needs more time to process data and process control is more complicated.

Speed and accuracy of an algorithm are key factors for the online visual inspection system. Hence, parallel algorithm is chosen to process pseudo-foreign fibers' images more suitably. Meanwhile, due to pseudo-foreign fibers have many kinds and colors, and pixels spread are extensive, so all of these specifications are fit for highly adaptability and flexibility of parallel algorithm. Here genetic algorithm is the most appropriate one. Using this algorithm to process threshold, it ensures segmentation's accuracy, and faster speed of this algorithm is more attractive.

2.4 Traditional Genetic Algorithm

Genetic algorithm is a bionic algorithm, the living beings in nature have undergo elimination and evolution continuously from the initial single-celled organisms to multicellular organism, and eventually human beings who have high intelligence and high adaptability of primates [8-9]. The process of evolution totally proved Darwin's "natural selection, survival of the fittest" theory [10- 12]. It told us the living beings who adapt to environment will have chance to breed, and others who can not accustom environment will be eliminated. Though iteration of dozens of generations, the quality of biological communities will improve [13]. Genetic algorithm used this principle to encode the parameter space, also do the steps of selection, crossover and mutation and other operations, then use the fitness function value the to evaluate images, the image information which is evaluated up to standard will be preserved into the next generation, the unqualified image information will be eliminated [14].

In addition, the genetic algorithm has many advantages for image segmentation of pseudo-foreign fibers. It is a probabilistic, iterative global search algorithm for groupment [15], the search space does not need restrictive assumptions as other image segmentation algorithms, such as: continuous, derivative exists, histogram single peak characteristics, etc. [16]; the searching is not blind, it gains on the best value step by step to finish the "evolution".

2.5 Improved Genetic Algorithm

As the traditional genetic algorithm has appeared "throwback" phenomenon in the process of choosing individual [17], in other words, offspring's fitness function value is less than the parent generation; meanwhile, in order to improve the processing speed of genetic algorithm, this paper did several improvements.

2.5.1 Improvement of Searching Range for Calculating Optimal Threshold

The genetic algorithm is a parallel global optimization algorithm, although the processing speed has obvious advantages than other algorithm, the selection of groupment is random, so it will waste some time. After tested a lot of samples, the distribution of the best segmentation threshold of the pseudo-foreign fibers images is concentrated from 120 to 200. In order to avoid the limitations of the tests and not rule out some special pseudo-foreign fibers which have special color, finally, the optimal threshold distribution is conformed from 100 to 220. This algorithm has reduced the range of searching, greatly decreasing the optimal threshold search time, the average processing time is tens of milliseconds, and the calculating speed in this stage was improved more than twice in average.

2.5.2 Combination of Elitist Strategy and Roulette Wheel Method

As the genetic algorithm is a random selection of individuals, it is possible that the individual whose fitness value is higher has not been selected, but the individual whose fitness value is lower is selected. This is "throwback" phenomenon. According to the principle of bionics analogy, a conclusion can be acquired that the individual whose fitness value is lower may produce the next generation whose fitness is not high, that means these individuals have poor quality. To overcome the drawbacks of these individuals, elitist strategy is led in to choose some individuals who have high fitness value to come into the next generation directly, and other individuals use roulette wheel method to select according to the share of individual fan-shaped area of the roulette selection probability.

2.5.3 Fitness Amendments Formula

In the initial generations of the genetic algorithm may produce some individuals whose fitness value are high, finally these individuals may eventually fill the entire group, which would make the crossover which is used to generate new individuals lose its meaning, even more importantly, the convergence of the algorithm may converge the part optimal value ahead of time, rather than the global optimal value, this is called "population premature" problem. Hence, Goldberg linear stretch formula is fixed to improve the fitness function, which reduces the differences of fitness value between individuals with higher fitness with other individuals; it can maintain population diversity effectively, balance maximum and average fitness value as the optimal fitness function. The amendments formulas are as follows:

$$H^*(s) = \frac{0.5H_{avg}(s)}{H_{\max}(s) - H_{avg}(s)} H(s) + \frac{H_{\max}(s) - 1.5H_{avg}(s)}{H_{\max}(s) - H_{avg}(s)} H_{avg}(s) . \quad (1)$$

Where $H^*(s)$ is amendments fitness function; $H_{avg}(s), H_{\max}(s)$ is average entropy of the histogram for the current generation of individuals, is the maximum entropy of histogram for the current generation of individuals; $H(s)$ is the entropy of histogram for the current generation of individuals. The algorithm model is:

$$S = \arg\max H(s) . \quad (2)$$

$$H(s) = H_A(s) + H_B(s) = \ln P(1-P) + \frac{H_s}{P} + \frac{H_T - H_s}{1-P} . \quad (3)$$

$$H_T = -\sum_{i=0}^{n-1} p_i \ln p_i . \quad (4)$$

$$H_S = -\sum_{i=0}^{s} p_i \ln p_i . \quad (5)$$

$$P = \sum_{i=0}^{s} p_i . \quad (6)$$

Where *S* is the segmentation threshold, $H(s)$ is the image histogram entropy which is also the fitness function, $H_A(s), H_B(s)$ is the total entropy for the target *A* and background *B*, H_T, H_S is the entropy of all pixels and s the following thresholds (including s), *P*, p_i is the probability of gray level for the following threshold values of *s* (including *s*) and the probability of every gray level . Threshold

S which can make $H(s)$ reach the maximum threshold is defined as the optimal threshold for image segmentation, $H(s)$ is also the best fitness value.

2.5.4 Determination of the Optimal Crossover Probability

Crossover is the main method of genetic algorithm to generate new individuals, it is similar to the homologous chromosome synapsis in biological evolution, in other words, a certain bit which comes from an individual's binary code exchanges with another individual corresponding bit to produce a new individual. Crossover probability value is very important. the value is the larger the better normally, the range between 0.4 to 1. Different targets, different values. If the value is too large, perfect pattern of group activities will be destroyed, the evolution of computing will be interrupted; if the value is too small, the speed of generating a new individual is too slow. Therefore, this paper improved Crossover probability value as follows: in the first 30 generations, the crossover probability is designed as 0.6, each generation has six individuals to executing crossover, other individuals only copied themselves directly into the next generation; in the post 70 generations, the crossover probability is designed as 0.8, because the latter generations' fitness value is relative concentration, this crossover probability can increase group diversity, each generation has eight individuals to executing crossover, and the remaining two individuals simply copied themselves directly into next generation.

2.6 Image Segmentation Steps on the basis of Improved Genetic Algorithm

The image segmentation steps on the basis of improved genetic algorithm are as follows:

(1) The establishment of the fitness function

According to formula (2) - (6).

(2) The amendment of the fitness function

According to formula (1).

(3) The establishment of populations and coding

Several individuals are selected randomly, then they are encoded binary code for the gray value of the individual, each binary code likes a chromosome, and these individuals' chromosomes are as the first generation. After calculating the fitness function value for each corresponding individual, the sum of group fitness value can be calculated, and then the fitness value of every individual also can be obtained, finally following the elite strategy and roulette wheel method, the individuals can be extracted and copied.

(4) Crossover

Which is followed the new crossover probability proposed in the above.

(5) Mutation

The process modeled the process of organism's gene mutation: a gene on chromosomal locus may mutate as its allele, it usually caused certain phenotype changes. Here this refers to each individual's binary code from 0 to 1 on certain part, or from 1 to 0. For converging to the maximum fitness value as quickly as possible, the mutation probability is set as follows in this paper: the first thirty generations is 0.02; from the thirty to fifty generation is 0.03; the latter fifty generations is 0.02. Then the new individuals are chosen to compose a new generation instead of the original individuals.

(6) Termination conditions

When the algorithm reaches the maximum generations, or after 10 generations the group's highest fitness value has not changed, the algorithm stops.

(7) Image segmentation

The group's highest fitness value is defined as the optimal threshold to perform image segmentation.

3 Results and Analysis

The paper processed twelve kinds of pseudo-foreign fibers' image, and compared segmentation effect and the processing time with the Otsu method and the traditional genetic algorithm. Partial results are shown in Fig. 1-3 and Table 1-2. It can be seen from the results:

(1) This segmentation algorithm can split the pseudo-foreign fibers precisely, and retain a good edge of pseudo-foreign fibers, the segmentation images are clear, which are shown in Fig. 3. As the pseudo-foreign fibers' images distributed unevenly, target has a wide range of gray scale, by using the Otsu method calculate the optimal threshold, the between-class variance becomes smaller, the probability of faulting the target into the background increases, take Fig. 1(a) as

an example, the bottom of the grass blade was mistaken into the white background, that means the pseudo-foreign fibers target are retained half-baked, Fig. 1(b), 1(c) also exist such situation. Traditional genetic algorithm has very nearly the same segmentation effect, and can split basically the pseudo-foreign fibers target, but the overall effect is not as good as the algorithm of this paper, and the effect is shown in Fig. 2.

(2) It can remove a lot of noise, including Gaussian noise produced by the optical components and impulse noise of environment. Three kinds of segmentation methods have similar anti-interference capabilities.

(3) Processing time of improved genetic algorithm is only tens of milliseconds, high processing speed. However, the traditional genetic algorithm and Otsu method need relatively long time, which are shown in Table 1.

(4) The optimal segmentation threshold distribution of improved genetic algorithm is between 100 to 220. Specific data is shown in Table 2. Table 2 shows that the optimal segmentation thresholds which were the following three segmentation methods calculated are within a reasonable range, that is the optimal segmentation threshold should not exceed the target ceiling's gray-scale 220; the optimal segmentation threshold by using Otsu method obtained is too large or too small, a larger fluctuation range, sometimes closes to the limit, the target image information has been lost, the optimal segmentation threshold by using the genetic algorithm has low threshold range, the target image has also been retained completely.

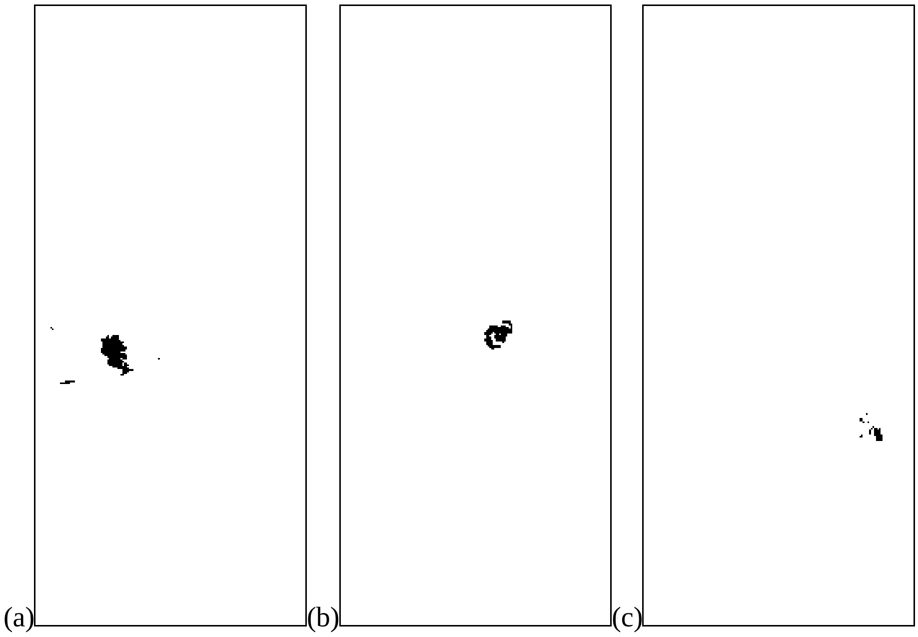

Fig.1. Segmentation results of the original gray images using the Otsu’s method.

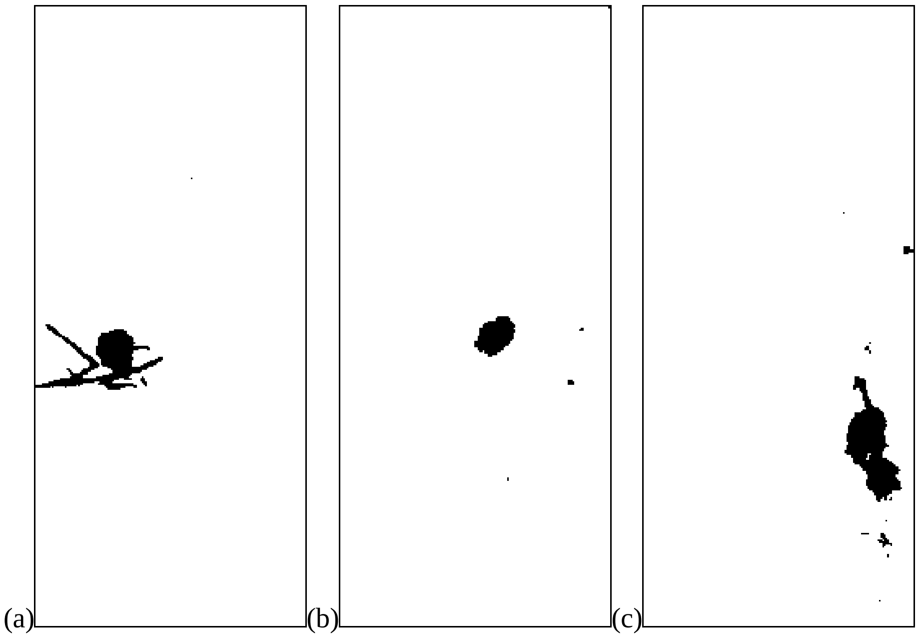

Fig.2. Segmentation results of the original gray images using the traditional genetic algorithm.

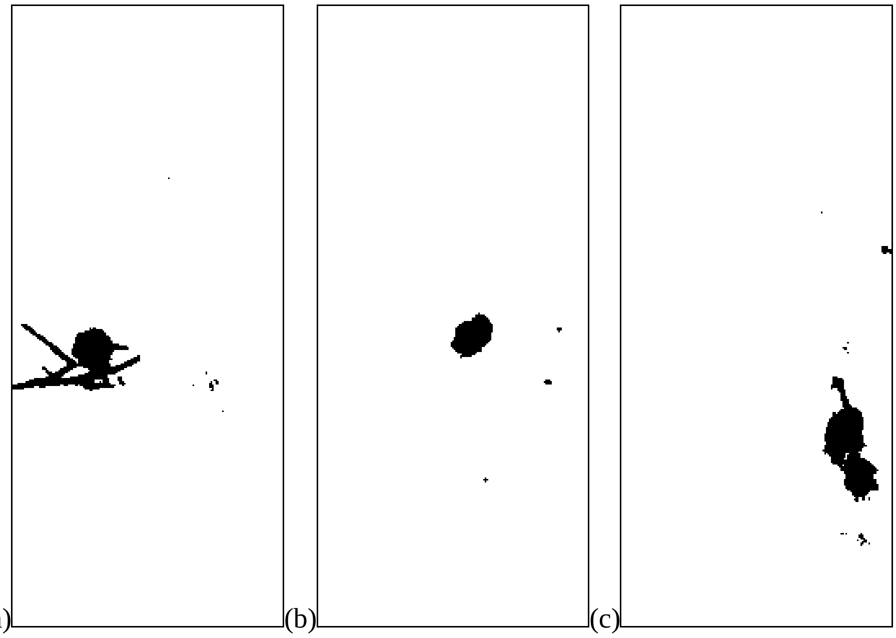
(a) (b) (c)

Fig.3. Segmentation results of the original gray images using the improved genetic algorithm.

Table 1. Image segmentation processing time comparison of pseudo-foreign fibers /ms.

	grass blade	cottonseed	stringy cotton
Otsu method	2942	406	437
traditional genetic algorithm	31	47	265
improved genetic algorithm	15	43	79

Table 2. The optimal segmentation threshold comparison of pseudo-foreign fibers in cotton.

image	the optimal segmentation threshold *S*		
	Otsu method	traditional genetic algorithm	improved genetic algorithm
grass blade	159	175	184
cottonseed	155	169	174

stringy cotton	200	179	178

4 Conclusion

By using digital imaging technology, this paper has processed the images of pseudo-foreign fibers in cotton, gray-scale characteristics of pseudo-foreign fibers and background are chosen as a starting point, presented a new image segmenting method on the basis of improved genetic algorithm. After the experiment contrast test, the method is verified that it can segment image precisely and clearly, also remove a lot of noise. Meanwhile, it avoided the target segmenting incompletely for a long time and other shortcomings of traditional methods. This new method reduced the searching range for calculating optimal threshold from 0~255 to 100~220. The calculating speed in this stage is improved more than twice in average. This paper proposed the fitness amendments formula, and also solved "premature", converging to global optimal solution difficultly issues of the traditional algorithm; this paper combined elitist strategy and roulette method to improve the algorithm. The results show that the algorithm has high-speed, accuracy, anti-interference and so on.

Acknowledgements

The authors thank National Natural Science Foundation of China (30971693) and Ministry of Education of People's Republic of China (NCET-09-0731), for their financial support, and China Cotton Machinery & Equipment Co., Ltd. for providing foreign fiber samples and technical support.

References

1. Ronghua,Ji., Daoliang,Li., Lairong,Chen., Wenzhu,Yang.: Classification and identification of foreign fibers in cotton on the basis of support vector machine [J]. Mathematical and Computer Modelling, 51: 1433--1437(2010)
2. Wenzhu,Yang., Sukui,Lu., Sile,Wang,, Daoliang,Li.:Fast recognition of foreign fibers in cotton lint using machine vision [J]. Mathematical and Computer Modelling, 54: 877--882 (2011)
3. Daoliang,Li., Wenzhu,Yang., Sile,Wang.: Classification of foreign fibers in cotton lint using

machine vision and multi-class support vector machine [J]. Computers and Electronics in Agriculture, 74:274--279(2010)
4. Jinming,Yang., Jianchang,Yuan., Ziqing,Wang.: Study on a foreign fiber detecting system with linear CCD [J]. Progress in Textile Science & Technology, 6:60--62(2009)
5. Baolin,Hu., Bin,Hu.: Difficulties of machine vision technology in the foreign fiber removed from the cotton [J]. China Cotton Processing, 5:31--33(2009)
6. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China. GB-1103-2007, 6:1--16(2007)
7. Wenzhu,Yang., Daoliang,Li., Xinhua,Wei., Yuguo,Kang., Futang,Li.: Toward Image Segmentation of Foreign Fibers in Lint [J]. Transactions of the Chinese Society for Agricultural Machinery, 40(3): 156--161(2009)
8. Xinquan,Lu., Ning,Li., Shifu,Chen.: Two-dimensional thresholding and genetic algorithms in image segmentations [J]. Computer Applications and Software, 12: 57--60(2001)
9. Lei,Wang., Tingzhi,Shen.: Two-dimensional entropy method based on genetic algorithm [J]. Journal of Beijing Institute of Technology, 11(2): 184--188(2002)
10. Zhengfeng,Li., Jie,Hou.: License plate image segmentation based on modified genetic algorithms [J]. Computer Engineering and Design, 26(9): 2455--2457(2005)
11. Yanghui,Wu., Naixin,Wang., Yingzhou,Liu.: Several populations competed genetic algorithm and its property analysis [J]. Journal of Northwest Sci-Tech University of Agriculture and Forestry, 33(4): 154--156(2005)
12. Lingyan,Wu., Tingzhi,Shen., Ziwen,Fang., Fei,Wang.: An image segmentation method using the entropy of histogram and genetic algorithms [J]. Acta armamentarii, 20(3): 255--258(1999)
13. Hongyan,Li., Liyuan,Sheng., Liangkuan,Chen., Gengqiang,Li.: Image Thresholding Segmentation Based on 2D Maximum Entropy Principle and Improved Genetic Algorithm [J]. Computer and Modernization, 2: 34--37(2007)
14. Zhenjie,Hou., Shoushi,Ma., Xichun,Pei,, Xin,Pan.: A Method of Marrow Cell Image Segmentation Based on GA [J]. Computer Engineering & Science, 28(10): 63--65(2006)
15. Yongjun,Li., Wei,Sun., Zaimei,Zhang.: Chromosome Biomedicine Image Processing Based on Genetic Algorithm [J]. Research and Exploration in Laboratory, 27 (5): 23--25(2008)
16. Yangting,Zheng., Hongqing,Zhu.: Image Segmentation Approach Based on Improved Genetic Algorithm [J]. Journal of Wuhan University of Technology, 28(3): 436--438(2004)
17. Goldberg,D: Genetic algorithms in search, optimization and Machine Learning [M]. USA: Addison -Wesley Publishing Company, 7 -10, 59 --308(1988)