

Foreign Fiber Image Segmentation Based on Maximum Entropy and Genetic Algorithm

Liping Chen, Xiangyang Chen, Sile Wang, Wenzhu Yang, Sukui Lu

School of Computer Science and Technology, Hebei University, Baoding, China
Email: mc_chenlp@sina.com

Received August 2015

Abstract

In machine-vision-based systems for detecting foreign fibers, due to the background of the cotton layer has the absolute advantage in the whole image, while the foreign fiber only account for a very small part, and what's more, the brightness and contrast of the image are all poor. Using the traditional image segmentation method, the segmentation results are very poor. By adopting the maximum entropy and genetic algorithm, the maximum entropy function was used as the fitness function of genetic algorithm. Through continuous optimization, the optimal segmentation threshold is determined. Experimental results prove that the image segmentation of this paper not only fast and accurate, but also has strong adaptability.

Keywords

Foreign Fibers, Image Segmentation, Maximum Entropy, Genetic Algorithm

1. Introduction

The foreign fibers in cotton refer to those non-cotton fibers and dyed fibers, such as polypropylene fiber silk, hemp, feathers, colored line, colored cloth, hairs and so on. Though very low content of foreign fibers in cotton, the presence of foreign fibers will seriously affect the quality of the final cotton textile products, as they may debase the strength of the yarn, they are not easy to be dyed, and this will lead to great economic loss for the cotton textile enterprises [1]. Therefore, fast and accurate measurement of foreign fibers in lint cotton enterprises is an urgent problem. Using machine vision technology to identify and measure the foreign fiber is an effective and feasible solution. The basis and key of the measurement of the foreign fiber content is the foreign fiber's classification, and the image segmentation is the premise and guarantee of the classification of the foreign fiber.

Image segmentation is one of the primary stages in image processing and machine vision system, and it is also the precondition of image analysis. The objective of image segmentation is to partition the image into meaningful connected-components to extract the features of objects [2]. According to the means of image segmentation, the image segmentation method can be divided into threshold method, boundary detection method, area method and so on. Among them, the threshold method is the earliest studied and used. At present, due to the characteristics of clear physical meaning, obviously effect, easily implementing and good real-time, the threshold method became one of the most commonly used image segmentation method in image analysis, image recognition and machine vision systems.

The current threshold segmentation method, according to the threshold selection criterion function types, can be divided into maximum entropy method [3], maximum inter-class variance (Otsu) method, cross entropy method [4], minimum error method, fuzzy entropy method and so on [5]. How to determine the optimal threshold is the key of the threshold segmentation method. Since the maximum entropy method does not require a priori knowledge, it is widely used in image segmentation. But the maximum entropy method is through exhaustive search of global optimal solution, it is very complex and slow. In recent years, due to the fast random search capability, genetic algorithm (GA) has been successfully applied in many fields, such as image processing, machine learning, parameter optimization and so on [6] [7]. In this paper, through combine GA with maximum entropy method, we realized the fast segmentation of the foreign fibers.

The remainder of this paper is organized as follows. Section 2 presents the maximum entropy method and the GA briefly. Section 3 provides the proposed algorithm in detail. Results and discussions are followed in Section 4. Section 5 declares the conclusion.

2. Methodology

2.1. Maximum Entropy

Maximum entropy algorithm is a segmentation method which based on the histogram of the image [8]. Let x be the gray level of the image (from 0 to L), $p(x)$ is the probability of the pixel which gray level is x in the image, then the entropy of the image is represented as:

$$H = -\sum_{x=1}^L p(x) \lg p(x) \quad (1)$$

Let the segmentation threshold is t , which will divide the image into two parts: target A and background B. Assuming that the gray values which less than the threshold of the pixels belong to the target A, the gray values which greater than or equal to the threshold of the pixels belong to the background B, then the probability of pixels in the image are related to target A and background B respectively is:

Target A:

$$p_a = \sum_{i=0}^{t-1} p_i \quad (2)$$

Background B:

$$p_b = \sum_{i=t}^{255} p_i \quad (3)$$

The entropy of the target region and the background region are respectively defined as:

$$H_a(t) = -\sum_{i=0}^{t-1} (p_i/p_a) \lg(p_i/p_a) \quad (4)$$

$$H_b(t) = -\sum_{i=t}^{255} (p_i/p_b) \lg(p_i/p_b) \quad (5)$$

The entropy of the whole image is defined as:

$$H(t) = H_a(t) + H_b(t) \quad (6)$$

The greater the value of function $H(t)$, the more information of the target and background represent, the more accurate the target and the background area. The global optimal threshold (let it be t^*) is the value which make $H(t)$ take the maximum value, that is, t^* can be expressed as:

$$t^* = \arg \max H(t) \quad (7)$$

2.2. Genetic Algorithm

Genetic algorithm (GA) is a parallel random search optimization method which simulates natural genetic mechanism and biological evolution theory. It is stochastic rather than exhaustive. It can obtain the global optimal

solution with large probability. In addition, GA does not require the evaluation function to be monotone, so GA is very suitable for real-time processing, which is not only easy to implement, but also has strong robustness. The basic idea of the algorithm is first encoding the problem to be solved into a string (named chromosome). Each chromosome is a solution to the being solved problem. Next, GA calculates the fitness value of each individual in a group of chromosomes, then chromosomes with higher fitness are selected for crossover and mutation operation and then produce a new generation with higher fitness. At last, the relative optimal solution is obtained by continuous iteration.

3. Foreign Fiber Image Segmentation Based on Maximum Entropy and Genetic Algorithm

Through analysis the cotton foreign fiber images, it is found that the background of the cotton layer has the absolute advantage in the whole image, while the target is very small (**Figures 2(a)-5(a)**). Furthermore, the brightness and contrast of the image are all poor. Using the traditional image segmentation method (e.g. Otsu), the segmentation results are very poor. Therefore, we adopt the maximum entropy threshold method in this paper. However, the essence of the maximum entropy method is to obtain the maximum value of the objective function (Equation (6)) in the gray space, the computation of the algorithm is very large, and it takes much more time. In order to reduce the complexity of the algorithm, we combine GA with the maximum entropy in the foreign fiber image segmentation. The maximum entropy function was used as the fitness function of GA. By continuous optimization, the optimal segmentation threshold is determined. The flow chart of the algorithm was shown as **Figure 1**.

3.1. Encoding and Design Fitness Function

Encoding a solution of a problem into a serial of genes, called chromosome, is very important when using GA. Various encoding methods have been proposed for particular problems to provide effective implementation of GA, such as binary encoding, real number encoding, integer or literal permutation encoding, or general data structure encoding [9]. In our research, we chose the binary encoding method. Since the gray level of the foreign fiber image has 256 gray level, so encoding the segmentation threshold with 8 binary string, and take the entropy function of Equation (6) as the fitness function.

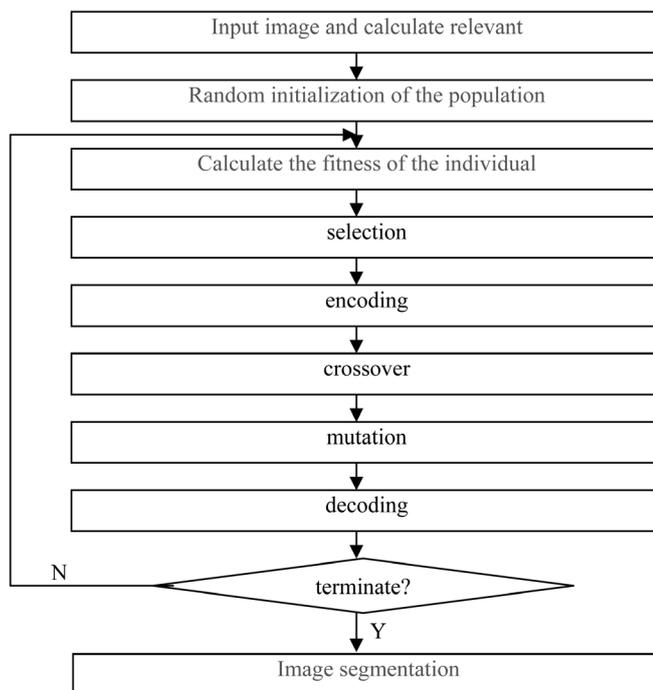


Figure 1. Flowchart of the proposed method.

3.2. Population Initialization

The initial population for evolution is created randomly. Individual in the initial population is equivalent to the candidate solution in the solution space. If the population size is too large, the computational complexity is high; if the scale is too small, the search space is limited, the search may stop at the immature stage. Therefore, the population number is set to 20, the maximum number of iterations is set to 100.

3.3. Selection Operation

The selection operation, also known as the copy operation, is to determine which individual is genetic and which individual is eliminated. The common methods include roulette wheel method, the elite preservation strategy and the order selection method. As the crossover and mutation operations are performed, the optimal solution can be easily lost in an intermediate step. So we adopted the method of 10% elite strategy and 90% roulette wheel. Individuals with the largest fitness directly into the next generation, which can guarantee that the algorithm converges to the global optimal solution; the remaining 90% of the individuals are selected according to the fitness proportionate. The larger the fitness of the individual, the greater the probability of being selected. The probability of the individual i is selected to the next generation group is:

$$P_i = F_i / \sum_{i=1}^n F_i \quad (8)$$

where n is the population size, F_i is the fitness of the individual i .

3.4. Crossover Operation

Commonly used crossover operations include single-point crossover, double-point crossover, multi-point crossover, etc. In our research, we chose the single-point crossover operator, the crossover probability is 0.8 and 0.6 respectively.

3.5. Mutation Operation

According to the mutation probability, the individual's value is replaced by some other gene values, which can increase the diversity of the population, and can also improve the local search ability of GA. Various mutation operations have been introduced by researchers, such as inversion mutation, neighbor exchange mutation, etc. We chose inversion mutation in our research.

3.6. Probability of Crossover and Mutation

The probabilities of crossover and mutation, denoted by P_c and P_m , affect seriously the search ability and convergence speed of GA [10]. For the standard GA, P_c and P_m are all fixed. But in the early stage of evolution, the individual difference is bigger, so we should increase P_c and reduce P_m in order to accelerate the evolution process and at the same time reducing the amount of calculation. In the later stage of evolution, the individual difference is small, so P_m should be reduced and P_c should be increased to avoid local optimal solution.

For the proposed algorithm, P_c is 0.8 during the first 50 generations of the evolution, so as to search the optimal solution as soon as possible. In the second part of the evolution, P_c is 0.6. The mutation probability P_m is calculated by:

$$P_m = \begin{cases} P_{m0} / \log_2 (g + 1) & f \geq \bar{f} \\ P_{m0} & f \leq \bar{f} \end{cases} \quad (9)$$

where P_{m0} is the initial mutation probability; f is the fitness of the parent chromosome; \bar{f} is the average fitness of the chromosomes in the current population [11].

3.7. Terminate Criterion

We set two terminate criteria in our GA: 1) maximum generation number. When the maximum number of iterations is reached, the algorithm terminates; 2) set the minimum number of iterations is 10, when the iteration

number is larger than the value, check the difference between the average fitness of the current population and the average fitness of the last generation whether less than a minimum value ε (here $\varepsilon = 0.0001$), once the condition is satisfied, the iteration terminates.

4. Results and Discussions

In order to verify the effectiveness of the paper's algorithm, we collected 60 foreign fiber images which include 5 foreign fibers such as hair, feather, hemp rope, polypropylene, cloth and so on. The image is 128×4096 size. The experiments are performed on over Matlab7.1 platform. We compared the results of the paper's method with the results of the Otsu's method. **Figures 2-5** show a part of the experimental results.

It can be seen from **Figures 2(a)-5(a)**, for the original image, the cotton layer background occupies the absolute advantage in the whole image, while the foreign fiber only accounts for a very small part, and the contrast of the image is very low. Therefore, in many cases, the target cannot be separated from the background by using the Otsu method (e.g. **Figures 2-4**). For **Figure 2(c)** and **Figure 3(c)**, although the target was segmented, but part of the background was divided into the target too. The foreign fiber of **Figure 4** was not extracted at all. So the reliability of the Otsu is very poor. In contrast, this paper's method works well in most cases. The extraction of the target is accurate and complete. **Table 1** lists the statistical segmentation results of 60 foreign fiber images with two different methods.

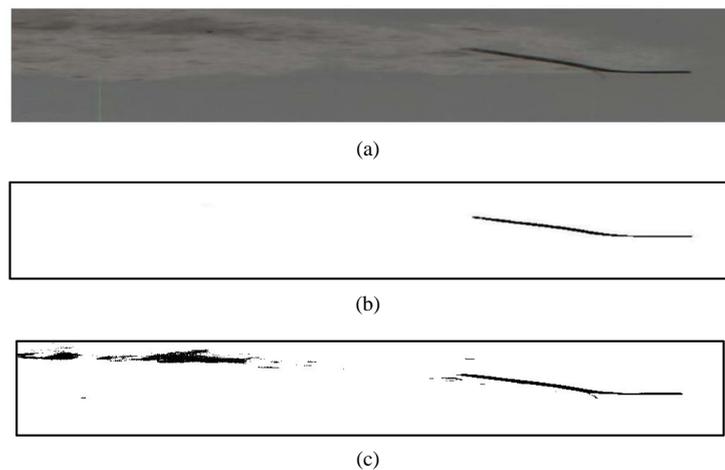


Figure 2. (a) Color image of polypropylene fiber; (b) The result of the paper's method; (c) The result of Otsu.

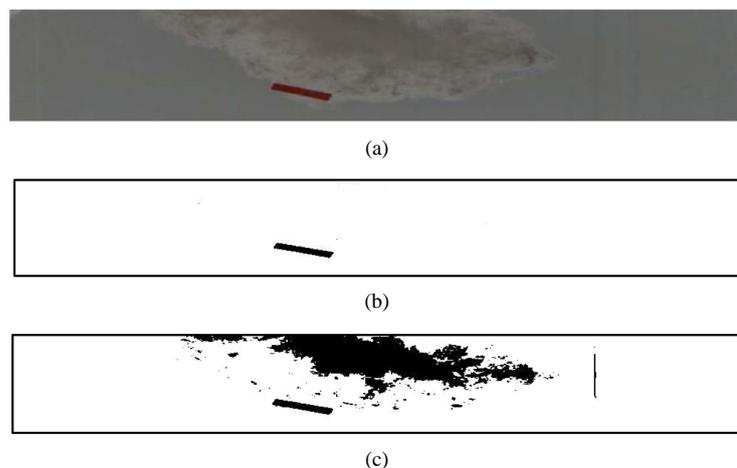


Figure 3. (a) Color image of cloth piece; (b) The result of the paper's method; (c) The result of Otsu.

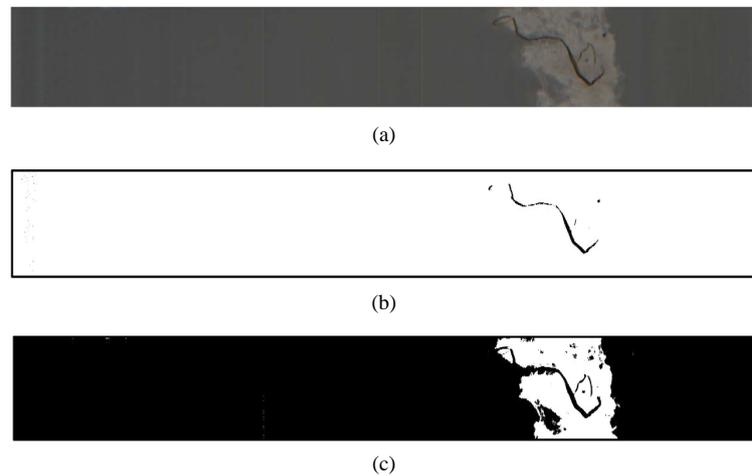


Figure 4. (a) Color image of hemp rope; (b) The result of the paper's method; (c) The result of Otsu.

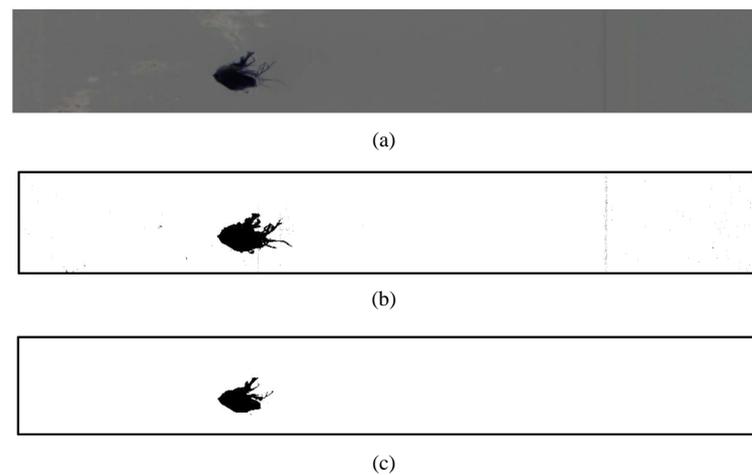


Figure 5. (a) Color image of feather; (b) The result of the paper's method; (c) The result of Otsu.

Table 1. Comparison of the segmentation results by two different methods.

Image name	Image number	Segmentation accuracy	
		The proposed method	Otsu
Hair	10	90%	20%
Feather	15	100%	100%
Polypropylene	10	100%	50%
Cloth	15	100%	60%
Hemp rope	10	100%	20%

The results shown in **Table 1** indicated that for different foreign fiber images, the segmentation accuracy of the proposed method is very high, it has strong adaptability. While for the Otsu method, the segmentation accuracy is higher when the area of the target is larger (such as the feather images, the segmentation accuracy is 100%). When the foreign fiber's area is small, the segmentation accuracy is very low (such as hair and twine images, the accuracy rate is only 20%). So, for low contrast images, the method of combining the maximum entropy and genetic algorithm can effectively improve the image segmentation accuracy.

5. Conclusions

Due to the low brightness, low contrast and small proportion of the target of the foreign fiber images, we adopted the image segmentation based on maximum entropy and genetic algorithm. Experimental results show that the adopted algorithm is more accurate than the traditional Otsu method, and the speed is fast and adaptable, which meets the real-time requirements of the foreign fiber detection system.

Next work: for the detection of white foreign fibers (such as plastic sheeting, plastic film, etc.), the accuracy of the segmentation algorithm is poor, there is still to be further improved.

Acknowledgements

The authors thank Hebei Natural Science Foundation (F2015201033), the Ministry of Science and Technology of the People's Republic of China (2013DFA11320), for their financial support.

References

- [1] Yang, M.H. (2014) Hazards and Prevention of Foreign Fiber in Cotton. *Jiangsu Textile*, **4**, 44-46. (In Chinese with English Abstract)
- [2] Zhang, Y.J. (2013) Image Engineering (II) Image Analysis. Tsinghua University Press, Beijing.
- [3] He, Z.J. and Wang, H.F. (2010) Chinese Word Sense Disambiguation Based on Maximum Entropy Model with Feature Selection. *Journal of Software*, **21**, 1287-1295. <http://dx.doi.org/10.3724/SP.J.1001.2010.03591>
- [4] Wu, Y.Q. and Zhang, X.J. (2011) Two-Dimensional Symmetric Cross-Entropy Image Thresholding. *Journal of Image and Graphics*, **16**, 1393-1401.
- [5] Wu, Y.Q., Meng, T.L. and Wu, S.H. (2015) Research Progress of Image Threshold Methods in Recent 20 Years (1994-2014). *Journal of Data Acquisition and Processing*, **30**, 1-23.
- [6] Ou, P. and He, D. (2011) 2-D Maximum Entropy Method of Image Segmentation Based on Genetic Algorithm. *Journal of Computer Simulation*, **1**, 294-297.
- [7] Guo, M.S. and Liu, B.H. (2008) 2-D Maximum Entropy Method in Image Segmentation Based on Chaos Genetic Algorithm. *Computer Technology and Development*, **18**, 101-104.
- [8] Cao, J.N. (2012) Review on Image Segmentation Based on Entropy. *Pattern Recognition and Artificial Intelligence*, **25**, 958-959.
- [9] Zhang, C.Q., Zheng, J.G. and Qian, J. (2011) Comparison of Coding Schemes for Genetic Algorithms. *Application Research of Computers*, **28**, 819-822.
- [10] Cao, D.Y. and Cheng, J.X. (2010) A Genetic Algorithm Based on Modified Selection Operator and Crossover Operator. *Computer Technology and Development*, **20**, 44-51.
- [11] Yang, W.Z., Li, D.L. and Zhu, L. (2010) An Improved Genetic Algorithm for Optimal Feature Subset Selection from Multi-Character Feature Set. *Journal of Agricultural Machinery*, **38**, 2733-2740.