Sensors & Transducers, Vol. 159, Issue 11, November 2013, pp. 32-38

www.sensorsportal.com

Sensors & Transducers

http://www.sensorsportal.com

Unsupervised Segmentation Method for Diseases of Soybean Color Image Based on Fuzzy Clustering

[*] Jiangsheng Gui, Li Hao, Shusen Sen, Wenshu Li, Yanfei Liu
College of Information, Zhejiang Sci-Tech University, Hangzhou 310018, China
[*] Tel: 086-571-86843320
[*] E-mail: dewgjs@126.com

Received: 28 August 2013 /Accepted: 25 October 2013 /Published: 30 November 2013

Abstract: The method of color image segmentation based on Fuzzy C-Means (FCM) clustering is simple, intuitive and is to be implemented. However, the clustering performance is affected by the center point of initialization and high computation and other issues. In this research, we propose a new color image unsupervised segmentation method based on fuzzy clustering. This method combines advantages of the fuzzy C-means algorithm and unsupervised clustering algorithm. Firstly, by gradually changing clusters c, and according to validity measurement, it can unsupervised search for optimal clusters c; then in order to achieve higher accuracy of clustering effect, the distance measurement scale was improved. In our experiments, this method was applied to color image segmentation for three kinds of soybean diseases. The results show that this method can more accurately segment the lesion area from the color image, and the segmentation processing of soybean disease is ideal, robustness, and have a high accuracy. *Copyright © 2013 IFSA.*

Keywords: Signal processing, FCM, Clustering, Diseases segmentation.

1. Introduction

As we all know, in many parts of world, soybeans are the main food of the people, and play an important role in easing the problem of hunger. But in recent years, due to some adverse factors such as natural disasters, soil erosion, fertilizer unreasonable lead to the occurrence of crop diseases, seriously affect soybean yield and quality in some areas. Diseases of field crops seriously threat to food security and the development of modern agriculture in China. Therefore, measure for scientific prevention and control of pests and diseases is important to ensure that the field crops have a good harvest. But the disease caused by the pathogen is often difficult to directly judge by human visual, and disease symptoms are influenced by a variety of factors, some potential information is covered with noise, these factors may be cause subjective misjudgment. Instead of human vision, computer vision can fast, accurate and effective in real-time diagnosis of disease information. With the development of image processing technology, it is possible to recognize crop diseases by image. Image recognition of crop disease includes image preprocessing, image segmentation, feature extraction and pattern recognition. Image segmentation is one of the key steps, and segmentation precision directly affects the reliability of feature extraction and the accuracy of pattern recognition. There are many methods currently used for image segmentation of crop diseases. Wang Shuwen proposed a watershed algorithm on cucumber leaves disease segmentation [1-2], but this algorithm is extremely sensitive to the noise in the image, and prone to split. Li Guanlin proposed a segmentation method based on k-means

Article number P_1534

hard Clustering Algorithm for color diseases image segmentation of grape [3-4], the algorithm requires constant sample classification adjustments, continuously calculate an adjusted new cluster centers, therefore, when the amount of data is very large, the time overhead of algorithm is very large. Yuan Yuan proposed crop disease leaves image segmentation based on Level set method [5-6], however, the Local Binary Fitting model for the segmentation of textured images cannot get good results. Kai Song put forward a method based on fuzzy c-means clustering to segment the corn diseases leaves [7-9], this method need to determine the number of clusters in advance and initialize the center point, slow calculation is also a problem. Camargo submit a method [10-11] to segment the crop diseases leaves, transform the RGB space to the HSV, I1I2I3 space, and extract H, I3a, I3b components, respectively threshold segmentation for the three components, but segmentation is not satisfactory.

In this research, according to the features of soybean leafs which infect gray spot, black spot, leaf spot, a segmentation method based on unsupervised fuzzy clustering was proposed. Unsupervised optimal fuzzy clustering [12-14] is a clustering algorithm based on fuzzy c-means algorithm. In the case of an unknown number of clustering, regardless of clustering specific shape (spherical or oval), density, size, scale, and using fuzzy super volume and separation density functions can assess the effectiveness of clustering, and automatically obtain optimize number of clusters c and cluster centers. So the speed of fuzzy clustering can be increased significantly and achieve segmentation of color image of soybean disease.

2. Method

This research mainly describes an algorithm for detection and identification of crop diseases through the analysis of color images of soybean leaves. This algorithm was divided into three steps:

1) Background segmentation: use Poisson matting technical to segment soybean disease leaves from background.

2) Preprocessing: the sharpening filtering technology was applied to the preprocessing of soybean leaf lesion image, the quality of lesion image has improved significantly.

3) Lesion segmentation: use unsupervised segmentation of fuzzy clustering from disease of the leaves out of the lesion area.

2.1. Image Set

All pictures in this research were obtained from Anhui Academy of agricultural sciences. They provided a number of pictures of typical soybean disease, for example: gray leaf spot, leaf spot, black spot and so on, and some of these pictures contain symptoms characteristics which are from the same strain of infection soybean disease at different stages. All the pictures is in JPEG format in this research.

After acquiring the image, in order to reduce the speed of program operation, it is necessary to compress size of 2592 × 944 to size of 800 × 600 image by the same proportion. With the purpose of improving the accuracy of segmentation, it is need to increase the contrast of the department with leaf disease and health, and the image was processed by the binarization and median filtering for binarization segmentation and directly after vector median filtering for color segmentation.

2.2. Fuzzy C-means Clustering Algorithm

Fuzzy C-means clustering (FCM) algorithm, which was proposed based on extending hard C-means (HCM) algorithm by Dunn [7, 15]. Bezked further promoted this work to an infinite cluster of vague objective function clustering, and proved the convergence of the algorithm [7, 16]. Fuzzy c-means (FCM) [17] is a method of clustering which allows one piece of data to belong to two or more clusters; it is based on minimization of the following objective function:

First, clustering loss function defined by Membership function can be written as

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} [u_{ij}]^b \parallel x_i - m_j \parallel^2, \tag{1}$$

where c is the number of the clusters, and other notations are described as follows.

x_i (i =1,2,…n) is the sample consisting of n samples collection, c is a predetermined number of categories, m_j (j = 1, 2, . . . , c) is the prototype of the j th cluster. If one pattern x_i belongs to the j th cluster, that means the distance between x_i and m_j is smallest. The exponent parameter b is used to control the influence of intermediate membership values on the objective function, and $1 < m < \infty$. u_{ij} is the fuzzy membership matrix. Where u_{ij} denotes the grade of membership of the i th pattern in the j th cluster. And it should satisfy the following there conditions:

1) $u_{ij} \in [0,1]\ (i = 1,2,...n; j = 1,2,...c)$

2) $\sum_{j=1}^{c} u_{ij} = 1\ \ (i = 1,2,...n)$

3) $0 < \sum_{j+1}^{c} u_{ij} < n\ \ (j = 1,2,...c)$

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers m_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\| x_i - m_j \|}{\| x_i - m_k \|}\right)^{\frac{2}{b-1}}} \tag{2}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}{}^{b} x_i}{\sum_{i=1}^{N} u_{ij}{}^{b}} \tag{3}$$

This iteration will stop when

$$\max_{ij}\{| u_{ij}{}^{(k+1)} - u_{ij}{}^{k} |\} < \varepsilon , \tag{4}$$

where ε is the termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

The algorithm is composed of the following steps:

1) Initialize U= [u_{ij}] matrix, U (0)

2) At k-step: calculate the centers vectors C (k) = [c_j] with U (k)

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}{}^{b} x_i}{\sum_{i=1}^{N} u_{ij}{}^{b}} \tag{5}$$

3) Update U (k), U (k+1)

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\| x_i - m_j \|}{\| x_i - m_k \|}\right)^{\frac{2}{b-1}}} \tag{6}$$

This iteration will stop when

$$\| U^{(k+1)} - U^{(k)} \| < \varepsilon \tag{7}$$

Otherwise return to step 2.

2.3. The Unsupervised Fuzzy Clustering Algorithm

Fuzzy C-means clustering (FCM) algorithm, which was proposed based on extending hard C-means (HCM) algorithm by Dunn [7, 12]. Bezked further promoted this work to an infinite cluster of vague objective function clustering, and proved the convergence of the algorithm [7, 13]. FCM has been widely applied in the field of image segmentation, but it still existed some problems as follows. Convergence to a local extremum; algorithm performance depends on the initial cluster centers; required to determine the number of clusters in advance and large amount of calculation. Unsupervised optimal fuzzy clustering is a clustering algorithm based on fuzzy c-means algorithm. In the case of an unknown number of clustering, regardless of clustering specific shape (spherical or oval), density, size, scale, and using fuzzy super volume and separation density functions can assess the effectiveness of clustering, and automatically obtain optimize number of clusters c and cluster centers. The basic idea of the present method is that gradually increase the number of clustering (less than a predetermined maximum number), iteratively, in the process of each iteration apply validation criteria to evaluate the effectiveness of clustering, so as to obtain optimal numbers of clusters and cluster centers. Its specific algorithms are as follows.

Step 1, select initial cluster centers, then set contrast coefficient m, maximum allowable error e (max) and the maximum number of cluster c(max).

Step 2, apply the fuzzy C-means algorithm to cluster and receive an initial cluster model, in which select the Euclidean distance as type of distance metric.

$$\mathrm{d}^2(x_{j,}v_i) = (x_j - v_i)^T(x_j - v_i), \tag{8}$$

where x_j (j = 1, 2,. . ., N) is the J m-dimensional feature vectors, v_i (i = 1, 2,. . ., c_0) is the center of the i th cluster.

Step 3, use the fuzzy c-means clustering algorithm, in which should change the distance function as exponential distance function, as the following.

$$\mathrm{d}^2(x_{j,}v_i) = \frac{1}{P_i}\sqrt{\det(F_i)}\exp[\frac{1}{2}(x_j - v_i)^T \tag{9}$$

in which P_i is for all members of the sample relative to the sum of the i th class of membership, namely:

$$P_i = \frac{1}{N}\sum_{j=1}^{N} u_{ji} \tag{10}$$

F_i is i th fuzzy covariance matrix, namely:

$$F_i = \frac{\sum_{j=1}^{N} u_{ji}(x_j - v_i)^T(x_j - v_i)}{\sum_{j=1}^{N} u_{ji}} \tag{11}$$

The step 4, calculate effectiveness of clustering, there are some parameters as following:

1) A fuzzy hypervolume standards:

$$v_{HV}(c) = \sum_{i=1}^{c} h_i \quad (12)$$

in which the hypervolume of i th class is h_i

$$h_i = \sqrt{\det(F_i)} \quad (13)$$

2) Classification density:

$$v_{PD}(c) = \frac{\sum_{i=1}^{c} s_i}{\sum_{i=1}^{c} h_i}, \quad (14)$$

where s_i is the sum of membership from all the good sample members (the members of the high degree of membership) in i th class:

$$s_i = \sum_{j=1}^{N} u_{ji} \quad (15)$$

$$\forall x_j \in \{x_j \mid u_{ji}(x_j - v_i)^T F_i^{-1}(x_j - v_i) < 1\}$$

normally, in order to avoid complicated calculation, used $u^{good} = 0.5$, therefore, the above equation becomes:

$$s_i = \sum_{j=1}^{N} u_{ji} \quad u_{ji} > 0.5 \quad (16)$$

3) Average density of separation:

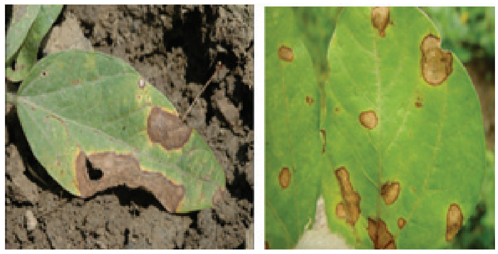
$$v_{APD}(c) = \frac{1}{c} v_{PD} \quad (17)$$

Step 5: If this cluster is less than a predetermined maximum number of clusters, the number of clusters plus 1 and repeat from step two, or stop the calculation and apply effective criteria to select the clusters number of average separation density maximum as the optimized clustering model.

3. Results

20 pictures set in this study are highlighting 3 typical soybean diseases (see Fig. 1). For the characteristics of the sample image of soybean diseases, the Poisson matting is applied to extract soybean leaves from complex background (see Fig. 2).

(a) gray spot (b) leaf spot



(c) black spot (d) gray spot (2)

Fig. 1. Original image.

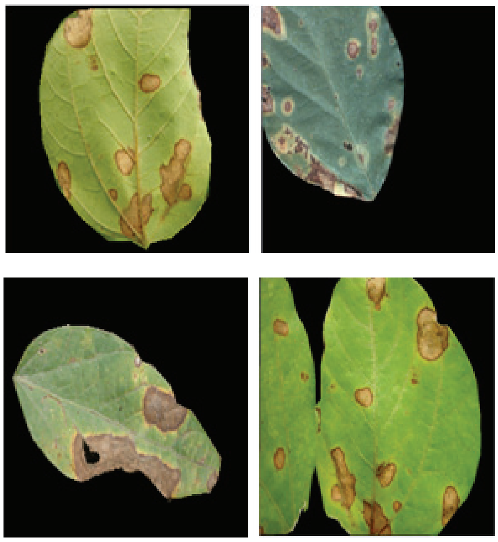


Fig. 2. Segmentation of leaf.

In this research, the input image is enhanced by using sharpening filter to make a blurry edge image and contour image in transmission and transformation process become clear, meanwhile the details become more apparent. Enhanced effects as shown in Fig. 3.

According to characteristics of sample image for soybean disease, in case study of gray spot, leaf spot and black spot disease leafs, unsupervised fuzzy clustering method is applied to segment the lesion part. In order to illustrate the advantages of the proposed method, here are the results of the three methods, segmentation results as shown Fig. 4.

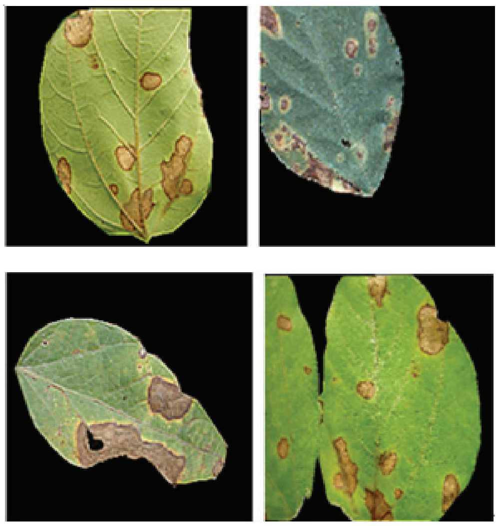


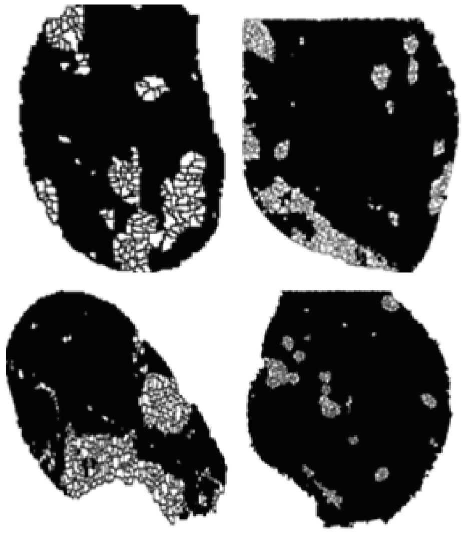
Fig. 3. Images after enhance transform.

From the graph, we can see in Fig. 4 (a) that watershed algorithm is sensitive to noise in image, so it is very easy to cause segmentation contour offset, and prone to over-segmentation. From Fig. 4 (b), it is show Mean shift (MS) algorithm [18-20] is an effective statistical iterative algorithm. Mean Shift algorithm are influenced by bandwidth the selected kernel function, the speed of Calculation is slow and the accuracy of segmentation is not high. Unsupervised fuzzy clustering method can preferably cluster the lesion part and normal part into two categories, the effect of segmentation is ideal.

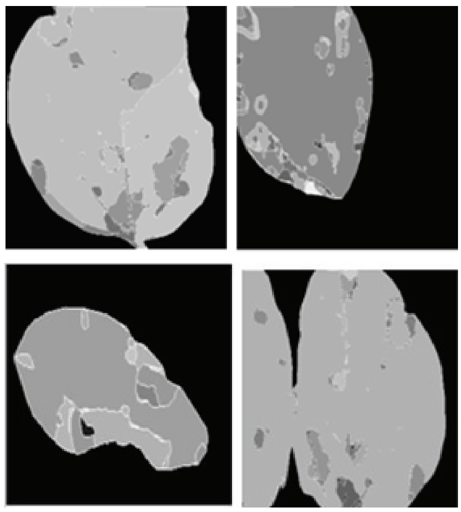
After disease image segmentation, there are often isolated points, burr and small cavities in the image. In order to reduce the noise effect on the subsequent identification, open and close operation in mathematical morphology was applied in this study to eliminate noise. The results were shown in Fig. 5. It can be seen that clustering segmentation result can be effectively corrected by mathematical morphology, at the same time eliminate the effect of leaf venation on segmentation of regions, improve the accuracy of segmentation.

4. Conclusions

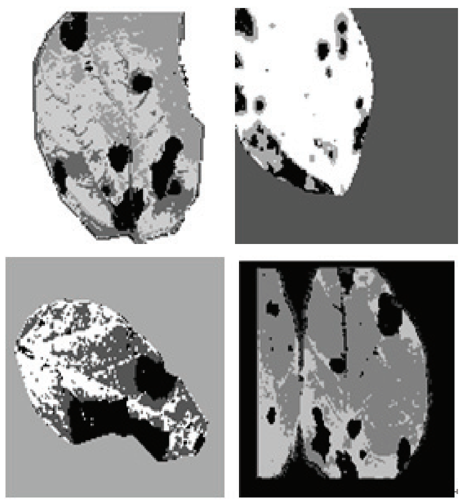
In this research, soybean black spot, gray leaf spot, leaf spot obvious symptoms of the color image were used as the research object, according to the features of three kinds of disease symptoms, a segmentation method based on unsupervised fuzzy clustering was proposed. Unsupervised optimal fuzzy clustering is a clustering algorithm based on fuzzy c-means algorithm. Using unsupervised optimal fuzzy clustering algorithm can get the optimal number of clusters, while clustering model than the clustering model by fuzzy C-means algorithm has been more reliable performance, and there is a obvious improvement in the clustering effect. Using this method for soybean disease for color image segmentation, mathematical morphology operations is applied to correct the clustering results and improve the accuracy of segmentation.



(a) Watershed algorithm segmentation result.



(b) Mean shift algorithm segmentation result.



(c) Proposed method segmentation result.

Fig. 4. Segmentation of spot.

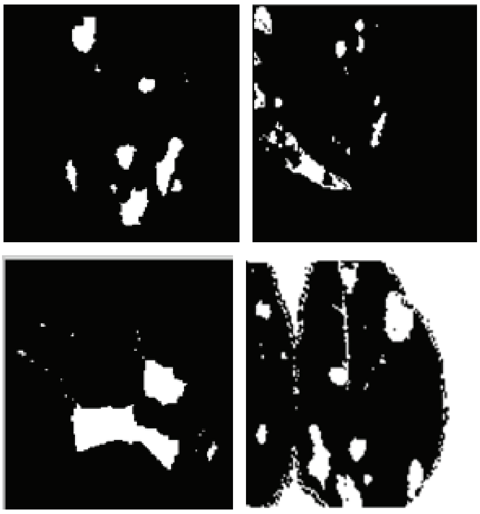


Fig. 5. Images after mathematical morphology operations.

The results show that this method can more accurately segment the lesion area from the color image, and segmentation processing of color image on soybean disease is ideal. Meanwhile the experiment results prove that the algorithm proposed in this paper is an efficient method for color images unsupervised segmentation.

Acknowledgements

The authors gratefully acknowledge the financial Support provided by the National Natural Science Foundation of China (Grant No. 61105035).

References

[1]. Wang Shuwen, Zhang Changli, Fang Junlong and Sekiteru, Research on the Identification of Cucumber Leaf Disease Based on Image Processing, *Heilongjiang Province Agricultural Engineering Academic Society*, 2011, pp. 173-179.

[2]. Ren Yugang, Zhang Jian, Li Miao and Yuan Yuan, Segmentation method for crop disease leaf images based on watershed algorithm, *Journal of Computer Applications*, 32, 3, 2012, pp. 752-755.

[3]. Li Guanlin, Ma Zhanhong, Huang Chong, Chi Yongwei and Wang Haiguang, Segmentation of color images of grape diseases using K_means clustering algorithm, *Transactions of the CSAE*, 26, Supp. 2, 2010, pp. 32-37.

[4]. Wang Shouzhi, He Dongjian, Li Wen, Wang Yanchun, Plant Leaf Disease Recognition Based on Kernel K-means Clustering Algorithm, *Transactions of the Chinese Society for Agricultural Machinery*, 40, 3, 2009, pp. 152-155.

[5]. Yuan Yuan, Li Miao, Qing Liang, Hu Xiuzhen and Zhang Wei, Segmentation method for crop disease leaf images with complex background, *Transactions of the CSAE*, 27, 2, 2011, pp. 208-212.

[6]. Chen Yuantao, Xu Weihong and Wu Jiaying, New colorful images segmentation algorithm based on level set, *Journal of Computer Applications*, 32, 3, 2012, pp. 749-751, 755.

[7]. Sompom Chuai-Aree, Chidchanok Lursinsap, Peraphon Sophatsathit, and Suchada Siripant, Fuzzy C-Mean: A statistical feature Classification of text and image segmentation method, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9, 6, December 2001, pp. 661-671.

[8]. Mao Hanping, Zhang Yancheng, Hu Bo, Segmentation of crop disease leaf images using fuzzy C-means clustering algorithm, *Transactions of the CSAE*, 2008, 24, 9, pp. 136-140.

[9]. Song Qingkun, Hao Min, Improved fuzzy C-means clustering algorithm, *Journal Harbin Univ. Sci. & Tech.*, 12, 4, 2007, pp. 8-10.

[10]. Camargo A., Smith J. S., Image pattern classification for the identification of disease causing agents in plants, *Computers and Electronics in Agriculture*, 66, 2009, pp. 121-125.

[11]. Camargo A., Smith J. S., An image-processing based algorithm to automatically identify plant disease visual symptoms, *Biosystems Engineering*, 102, 1, 2009, pp. 9-21.

[12]. Gao Yongqing, Chen Zhihong, Huang Heling, Ni Zhao Bin and Li Ting, Unsupervised optimal fuzzy clustering algorithm based on FCM, *Information Technology,* 7, 2009, pp. 69-71.

[13]. Xuejian Xiong, Kap Luk Chan, Towards An Unsupervised Optimal Fuzzy Clustering Algorithm for Image Database Organization, Pattern Recognition, in *Proceedings of the 15th International Conference*, 2000, pp. 897-900.

[14]. Yimin Hou, Xiangmin Lun, Wei Meng, Tao Liu, Xiaoli Sun, Unsupervised Segmentation Method for Color Image Based on MRF, Computational Intelligence and Natural Computing, in *Proceedings of the CINC'09 International Conference,* 2009, pp. 174-177.

[15]. Wang Hao, Wang Xiuyou and Chen Yun, New hybrid fuzzy C-means clustering algorithm, *Computer Engineering and Design*, 29, 4, 2008, pp. 917-922.

[16]. Cheng H. D., Jiang X. H. and Sun Y., Color Image Segmentation: Advances and Prospects, *Pattern Recognition*, 34, 12, 2001, pp. 259-228.

[17]. Zhu Xiaoyu, Color Image Segmentation Algorithm Based on FCM Clustering, Master's Thesis, *Jilin University,* 12, 2011.

[18]. Comanicu D., Meer P., Mean shift: a robust approach toward feature space analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 5, 2002, pp. 603-619.

[19]. Wang Zhaoying, Li Guang, Hou Zhiqiang, Color image segmentation by fusion of mean shift and region saliency, *Computer Engineering and Applications*, 46, 28, 2010, pp. 181-184.

[20]. Cao Xiaolin, Application of Mean Shift algorithm in the segmentation of color image, *Software Guide*, 10, 11, 2011, pp. 79-81.