

Plant disease leaf image segmentation based on superpixel clustering and EM algorithm

Shanwen Zhang¹  · Zhuhong You¹ · Xiaowei Wu¹

Received: 15 January 2017 / Accepted: 13 June 2017
© The Natural Computing Applications Forum 2017

Abstract Plant disease leaf image segmentation plays an important role in the plant disease detection through leaf symptoms. A novel segmentation method of plant disease leaf image is proposed based on a hybrid clustering. The whole color leaf image is firstly divided into a number of compact and nearly uniform superpixels by superpixel clustering, which can provide useful clustering cues to guide image segmentation to accelerate the convergence speed of the expectation maximization (EM) algorithm, and then, the lesion pixels are quickly and accurately segmented from each superpixel by EM algorithm. The experimental results and the comparison results with similar approaches demonstrate that the proposed method is effective and has high practical value for plant disease detection.

Keywords Plant disease leaf image segmentation · Plant disease detection · Superpixel clustering · EM algorithm

1 Introduction

In agriculture field, there are various plant diseases, which seriously reduce the quantity and quality of agriculture products. Automatic detection of plant diseases through plant leaf symptoms applied in leaves is a relatively affordable and more effective solution as it can reduce a large work of monitoring in big plant fields. In the plant disease detection, leaf image segmentation is quite

important, which directly affects the reliability of feature extraction and disease recognition accuracy. Recently, many methods have been proposed for plant disease leaf image segmentation based on statistical pattern recognition, *K*-means clustering, fuzzy *C*-means clustering, Otsu, level set, EM algorithm, etc. [1, 2]. Yuan et al. [3] proposed a segmentation method for crop disease leaf with complex background, in which an optimal threshold value for segmentation can be obtained using weighted Parzen window. Arivazhagan et al. [4] proposed a software solution for automatic detection and classification of plant leaf diseases, in which the green pixels are masked and removed using specific threshold value followed by segmentation process, and the texture statistics are computed for the useful segmentation. Gui et al. [5] presented a soybean leaf disease detection system using color leaf image with complex background, in which according to the *K*-means algorithm and combined experience threshold, salient regions can be segmented from soybean leaf disease images based on salient maps. Hiary et al. [6] presented an algorithm for leaf disease image segmentation. In the algorithm, most green colored pixels are firstly identified, and then, these pixels are masked based on specific threshold values that are computed using Otsu's method, and finally the pixels with zero values of red, green and blue and the pixels on the boundaries of the infected cluster were completely removed. Yugang et al. [7] proposed a watershed algorithm for crop disease leaf segmentation. Kaur et al. [8] proposed a segmentation method based on *K*-means clustering algorithm for grape disease leaf image segmentation. Chaudhary et al. [9] proposed an algorithm for disease spot segmentation using image processing techniques in plant leaf, in which they compared the effect of CIELAB, HSI and YCbCr color spaces in the process of disease spot detection. Hanping [10] proposed a cotton

✉ Shanwen Zhang
zhangshanwen@xijing.edu.cn

¹ Department of Information Engineering, XiJing University, Xi'an 710123, China

disease leaf segmentation method based on fuzzy C -means clustering. Meunkaewjinda et al. [11] segmented grape leaf disease image by unsupervised optimal fuzzy C -means clustering algorithm. Qin et al. [12] investigated twelve lesion segmentation methods integrated with clustering algorithms for alfalfa leaf disease recognition. Mohammad et al. [13] compared the thresholding, watershed, edge detection, fuzzy C -means clustering and K -means clustering for plant disease detection. They pointed that fuzzy clustering algorithm is suitable for overlapping clustering task.

From the above analysis, it is known that disease leaf image segmentation is always a difficult and challenging problem because of the fuzziness and complexity of disease leaf image and the overlapping of the different regions. Most disease leaf segmentation algorithms utilize a fixed threshold or criterion to distinguish leaf image spot by gray-level differences between the normal pixels, spot pixels and background pixels. But in practice, the regions of these pixels in a disease leaf image are generally fuzzy and uncertain, and the color of normal region and spot region is also uneven and unclear, and the gray histogram of disease leaf image is always overlapping, as shown in Fig. 1. In Fig. 1c, it is found that there is no obvious segmentation valley between normal part and spot part in the image. So, lesion image cannot be effectively segmented by a fixed threshold by histogram clustering, Otsu and level-set algorithms [1, 2], which is shown in Fig. 1d–f.

Figure 1g shows that randomly selecting K clustering centers results in the poor segmentation by K -means clustering.

In some iteration segmentation algorithms, such as K -means clustering and fuzzy C -means clustering, the iteration processing is conducted in the whole disease leaf image. The iterative processing of the whole image will involve a huge amount of data. For example, given a color disease leaf with size of 240×320 , there are $240 \times 320 \times 3 = 230,400$ pixel points in RGB space. Dealing with so large data at the same time will seriously reduce the iteration speed. In some local segmentation methods, such as salient segmentation [5], the whole leaf image is divided into a number of even square grid sub-regions, as shown in Fig. 2b. Although the rigid even dividing is simple, from Fig. 2b, we can see that the spot pixels and normal pixels are grouped arbitrarily into a grid, which is unreasonable.

Superpixel clustering is becoming increasingly popular in computer vision image processing [14]. It is a process that neighboring pixels with the same or similar features in color, brightness and texture in an image are grouped into several perceptually meaningful homogeneous regions, which can be used to reduce the complexity of images from hundreds of thousands of pixels to only a number of superpixels, and can be used to replace the even dividing [5], as shown in Fig. 2c with 50 superpixels. In Fig. 2c, it is seen that although superpixels look quite irregular, a leaf

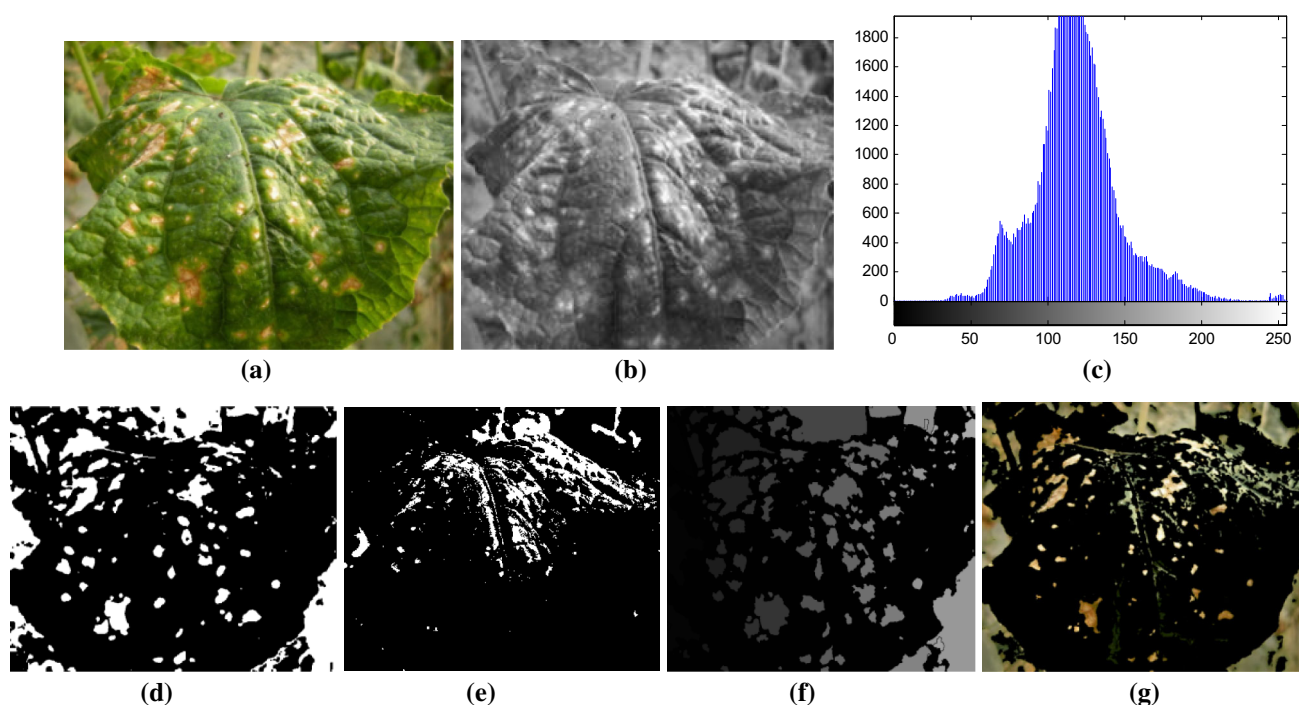


Fig. 1 Plant disease leaf image segmentation. **a** Original cucumber leaf image, **b** grayscale image, **c** grayscale histogram, **d** histogram clustering, **e** Otsu, **f** level set, **g** K -means clustering

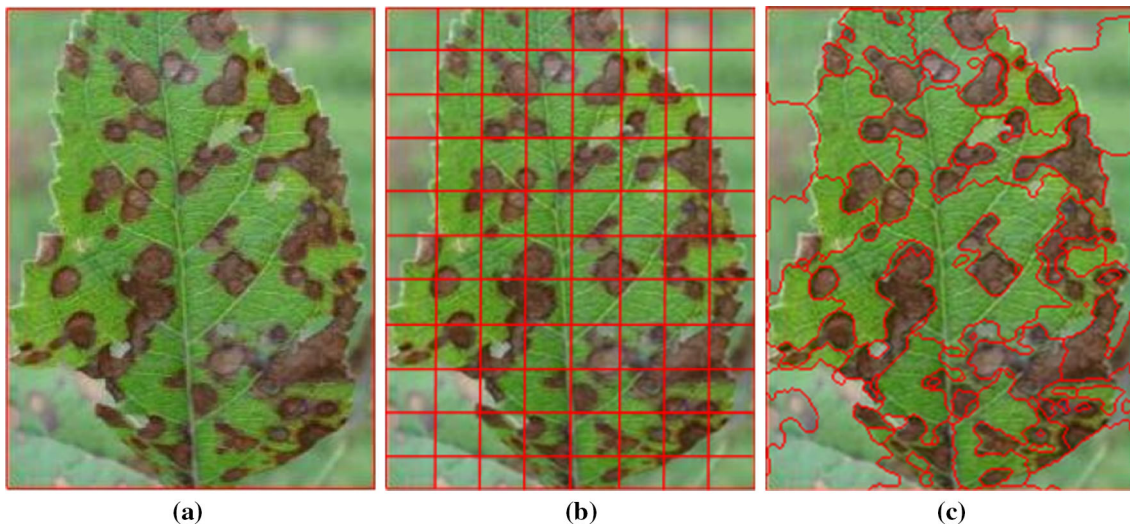


Fig. 2 Plant disease leaf image dividing. **a** Original disease leaf, **b** grid dividing, **c** superpixels

image is divided purposefully clustered. All pixels in the same superpixel are quite similar, and specially, the spot pixels can be grouped into several superpixels. There are many approaches to generating superpixels, each with its own advantages and drawbacks that may be better suited to a particular application [15–17]. These approaches can be broadly categorized as either graph-based or gradient ascent methods. Below, we review popular superpixel methods for each of these categories, including some that were not originally designed specifically to generate superpixels.

In disease leaf segmentation, superpixel clustering, as a preprocessing, can provide a compact representation of an original color disease leaf image. Superpixels of disease leaf image carry more information than individual pixels, which have great improvement in computational efficiency, running time and memory cost, because hundreds of thousands of pixels are grouped into only a few hundreds of superpixels [14]. The expectation maximization (EM) algorithm is frequently used for color image segmentation [18]. Although EM is effective for color image segmentation, it is difficult to determine its initial parameters. Specially, if its initial parameters are randomly set for whole image, it will take too much time to converge, which greatly affects its speed and constraints its practical applications. If we implement EM algorithm on each superpixel instead of on whole image, the initialize parameters can be easily estimated, because the pixels in a superpixel are similar and compact. In the paper, based on superpixels and EM algorithm, a novel method of plant disease leaf image segmentation is proposed. In the method, the superpixel clustering is employed to provide cluster cues to guide the lesion segmentation. EM contributes to the lesion segmentation by improving the

parameters of the Gaussian mixture density function on the basis of the maximum likelihood criterion.

The highlights are listed as follows: (1) superpixel clustering is applied to dividing the original color disease leaf image into a few hundreds of small compact regions; (2) EM algorithm can quickly converge by the guide of superpixel clustering, and the computational cost is greatly reduced; (3) the segmentation result is better and more robust to noise, since superpixels of area less than a given number of pixels are eliminated; (4) the color disease leaf images can be fast and accurately segmented by EM.

The rest of this paper is arranged as follows: Section 2 introduces superpixel clustering and EM algorithm. A segmentation method of plant disease leaf image is proposed by combining superpixels and EM algorithm in Sect. 3. The experimental results are presented and analyzed in Sect. 4. Finally, the conclusions and the future work are given in Sect. 5.

2 Related works

Superpixels can compactly represent disease leaf image and keep the neighbor relationship of pixels of image. EM algorithm is an efficient iterative algorithm for finding the maximum likelihood estimation. Superpixel clustering and EM algorithm are often applied to color image segmentation.

2.1 Simple linear iterative clustering (SLIC)

Simple linear iterative clustering (SLIC) is widely applied to superpixel clustering due to its simplicity and practicality [19]. It divides an image into approximately several

smaller regions, which are sufficient to preserve well the different boundaries in the image. The only parameter in SLIC is K , a desired number of approximately equal-sized superpixels. SLIC utilizes a distance metric to determine the closest cluster center for each pixel and then iteratively assigns pixels to the closest cluster center and updates the locations of the cluster centers. In CIELAB color space, the Lab color space describes mathematically all perceivable colors in the three dimensions: L for lightness and a and b for the color opponents green–red and blue–yellow. Unlike the RGB and CMYK color models, Lab color is designed to approximate human vision. In clustering, color image is converted to CIELAB color space, which provides a perceptually uniform color space. SLIC performs a 5-D space ($L^*a^*b^*x$ y) clustering by K -means guideline, where $L^*a^*b^*$ are components of CIELAB color space and x and y are the pixel coordinates in the image [20].

Given an image with n pixels and the number of superpixels K , the approximate size of each superpixel is K/n and the grid interval is about $S = \sqrt{K/n}$. Let K superpixel cluster centers be $C_k = (L_k, a_k, b_k, x_k, y_k)$, $k = 1, 2, \dots, K$, and $(L_1, a_1, b_1, x_1, y_1)$ and $(L_2, a_2, b_2, x_2, y_2)$ are two random cluster centers. Three distance measures in SLIC are defined in CIELAB color space as follows:

$$\begin{cases} d_{Lab} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \\ d_{xy} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \\ D_s = d_{Lab} + (\beta/S) \cdot d_{xy} \end{cases} \quad (1)$$

where D_s is the sum of $L^*a^*b^*$ distance d_{Lab} and xy plane distance d_{xy} normalized by the grid interval S , $1 \leq \beta \leq 20$ is introduced in D_s allowing to control the compactness of superpixels. The greater β , the more spatial proximity is emphasized and the more cluster compact.

SLIC is summarized as follows,

Step 1 Initialize cluster centers $C_k = [L_k, a_k, b_k, x_k, y_k]^T$ by sampling pixels at regular grid S .

Step 2 Perturb cluster centers in a 3×3 neighborhood to the lowest gradient position. Image gradients are computed as

$$\begin{aligned} G(x, y) = & \|I(x + 1, y) - I(x - 1, y)\|^2 \\ & + \|I(x, y + 1) - I(x, y - 1)\|^2, \end{aligned} \quad (2)$$

where $I(x, y)$ is the lab vector corresponding to the (x, y) pixel.

Step 3 For each cluster center C_k , assign each pixel with the nearest cluster center according to D_s , considering only the centers whose search region of $2S \times 2S$ pixels overlaps its location.

Step 4 Adjust the cluster centers to be the average vector of all the pixels belonging to the cluster.

Step 5 Return to Step 3. Compute new cluster centers and residual error E , i.e., the distance between previous centers and recomputed centers until $E \leq$ threshold.

2.2 EM algorithm

EM algorithm consists of E -step and M -step [21]. In E -step, a function is created for the expectation of the log-likelihood evaluation by the current estimate for its parameters; in M -step, the parameters are calculated by maximizing the expected log-likelihood found in the E -step. The estimated parameters are used to determine the latent variable distribution in the next E -step. In disease leaf image segmentation, EM searches the leaf image pixel clusters by determining the parameters of the mixture Gaussian density function that fits a given pixel set and assigns the pixels partially to different clusters instead of assigning it to only one cluster, and each cluster is modeled by a probabilistic distribution [22]. In general, for each superpixel, the region is small, and the pixel color is similar and compact, and the color change is not too large. So all pixels in each superpixel can be divided into healthy case and disease case, and they approximately obey the weighted average of mixture two-Gaussian distribution with two Gaussian densities.

Given a color disease leaf image with n pixels observed in $L^*a^*b^*$ color space, $V_i = [v_i^L, v_i^a, v_i^b](i = 1, 2, \dots, n)$ is the i th pixel. The mixture two-Gaussian density function can be expressed as

$$f(V_i; \aleph) = \sum_{q=1}^2 \alpha_q \cdot f_q(V_i; \mu_q, \Sigma_q) \quad (3)$$

where $\aleph = [\alpha_1, \alpha_2, \mu_1, \mu_2, \Sigma_1, \Sigma_2]$ is a parameter vector that needs to be estimated, $f_q(V_i; \mu_q, \Sigma_q)$ is the q th ($q = 1, 2$) mixture Gaussian distribution function with a mean vector $\mu_q \in R^3$ and a covariance matrix Σ_q , and α_q is the proportion of the q th normal density in the mixture, and $\alpha_1 + \alpha_2 = 1$. $f_q(V_i; \mu_q, \Sigma_q)$ is denoted as,

$$f_q(V_i; \mu_q, \Sigma_q) = \frac{1}{\sqrt{(2\pi)^3 |\Sigma_q|}} \exp \left[-\frac{1}{2} (V_i - \mu_q)^T \Sigma_q^{-1} (V_i - \mu_q) \right]. \quad (4)$$

EM algorithm is iteratively carried out with the initial parameter vector $\aleph^0 = [\alpha_1^0, \alpha_2^0, \mu_1^0, \mu_2^0, \Sigma_1^0, \Sigma_2^0]$ by repeating to update the parameter vector. Assuming $\aleph^t = [\alpha_1^t, \alpha_2^t, \mu_1^t, \mu_2^t, \Sigma_1^t, \Sigma_2^t]$ is the estimated parameter vector of the t -th iteration, \aleph^{t+1} of the $(t + 1)$ -th iteration can be estimated by,

$$\mu_q^{t+1} = \sum_{i=1}^n E_{iq}^t \cdot V_i / \sum_{i=1}^n E_{iq}^t \tag{5}$$

$$\Sigma_q^{t+1} = \sum_{i=1}^n E_{iq}^t \cdot (V_i - \mu_q^t)^T (V_i - \mu_q^t) / \sum_{i=1}^n E_{iq}^t \tag{6}$$

$$\alpha_q^{t+1} = \frac{1}{n} \sum_{i=1}^n E_{iq}^t \tag{7}$$

where $E_{iq}^t = \frac{\alpha_q^t f_q(V_i; \mu_q^t, \Sigma_q^t)}{\sum_{q=1}^2 \alpha_q^t f_q(V_i; \mu_q^t, \Sigma_q^t)}$, $q = 1, 2$.

Constantly repeating Eqs. (5)–(7) until the difference $|\aleph^{t+1} - \aleph^t|$ is less than a given threshold λ , and the final estimate \aleph^{EM} is obtained.

EM algorithm can be implemented by a mixture two-Gaussian model to get the probability of each pixel point belonging to each cluster, constantly update the estimated parameters and then gradually modify the center value of each cluster. Finally, the disease leaf image can be segmented by the classical maximum likelihood algorithm using \aleph^{EM} .

3 Plant disease leaf segmentation method

In color disease leaf image segmentation, superpixel clustering, as a preprocessing process, is used to provide a compact representation of the disease leaf image and further provides the initial parameters of the mixture two-Gaussian distribution for EM algorithm. The parameters can be used to describe the statistical characteristics of the healthy and disease pixels in the disease leaf image. By combining superpixel clustering and EM algorithm, a segmentation algorithm of plant disease leaf image, namely superpixels + EM, is summarized as follows.

1. Convert RGB color disease leaf image into $L^*a^*b^*$ color space.
2. Implement SLIC and obtain K superpixels, noted as I_1, I_2, \dots, I_K , where K can be determined experimentally.
3. Segment the lesion image from each superpixel by the following process.
 - (a) For the k th superpixel $I_k (k = 1, 2, \dots, K)$, estimate the initial parameters of EM algorithm: Suppose $V_{jk} = [v_{jk}^L, v_{jk}^a, v_{jk}^b] = [v_{1jk}, v_{2jk}, v_{3jk}]$, where v_{ijk} is the pixel value of the i th component of $L^*a^*b^*$ of the j th pixel in $I_k, i = 1, 2, 3; j = 1, 2, \dots, n_k, n_k$ is the pixel number of I_k , the initial parameters of EM algorithm can be estimated as

$$\begin{cases} \Sigma_{1k}^0 = \Sigma_{2k}^0 = E, \alpha_{1k}^0 = \alpha_{2k}^0 = 1/2 \\ \mu_{1k}^0 = [b_{11k}^0, b_{12k}^0, b_{13k}^0] \\ \mu_{2k}^0 = [b_{21k}^0, b_{22k}^0, b_{23k}^0], \end{cases} \tag{8}$$

where E is a 3×3 unit matrix, $b_{1ik}^0 = \frac{1}{n'_{ik}} \sum_{j=1}^{n_k} v_{ijk}$ and $v_{ijk} \leq \text{Mean}_{ik}$, $b_{2ik}^0 = \frac{1}{n_k - n'_{ik}} \sum_{j=1}^{n_k} v_{ijk}$ and $v_{ijk} > \text{Mean}_{ik}$, n'_{ik} is the pixel number of $v_{ijk} \leq \text{Mean}_{ik}$ in I_k , $\text{Mean}_{ik} = \frac{\text{Max}_{ik} + \text{Min}_{ik}}{2}$, $\text{Max}_{ik} = \max\{v_{ijk}, j = 1, 2, \dots, n_j\}$ and $\text{Min}_{ik} = \min\{v_{ijk}, j = 1, 2, \dots, n_j\}$.

- (b) Calculate the final parameter estimate

$$\aleph^{EM} = \{\alpha_{1k}^{EM}, \alpha_{2k}^{EM}, \mu_{1k}^{EM}, \mu_{2k}^{EM}, \Sigma_{1k}^{EM}, \Sigma_{2k}^{EM}\}$$

by EM algorithm.

- (c) Determine the label of each pixel in I_k :

The estimated parameters of EM algorithm can be used to divide each pixel into the healthy and disease regions, namely the lesion region segmentation. The j th pixel V_{jk} is labeled lab_j by

$$lab_j = \max_q \frac{\exp[-\frac{1}{2}(V_{jk} - \mu_{qk}^{EM})^T \cdot (\Sigma_{qk}^{EM})^{-1} \cdot (V_{jk} - \mu_{qk}^{EM})]}{|\Sigma_{qk}^{EM}|^{-1/2}} \tag{9}$$

where $q = 1$ means V_{jk} is healthy pixel; $q = 2$ means V_{jk} is lesion pixel [19].

In SLIC, the superpixel number of superpixels K influences computational complexity and boundary integrity. The search region of each K -means clustering center is inversely proportional to K . When K increases, SLIC will quickly converge to the global optimal. In other words, the larger K is, the more accurate edge segmentation is. However, if K is exceedingly large, it will lead to serious over-segmentation problem. The computational complexity to generate superpixels by SLLC differs with K . The practical computational complexity of SLIC is $O(nKN)$, where N is the number of iteration and is irrespective of K . Therefore, the reasonable selection of superpixel number is an important factor in SLIC algorithm.

4 Experiments and analysis

In order to examine and verify the performance of superpixels + EM, a number of experiments are conducted on a cucumber disease leaf image database, in a mixture platform of MATLAB 7.0 and under Windows XP and 120 G hard disk. All color disease cucumber leaves in the database were collected with digital camera in the agricultural

demonstration zone of Northwest Agriculture and Forestry University in China. The number of superpixels is set as 150; the threshold in EM algorithm is set as 0.01. Figure 3a shows four diseased leaf images, where two images with complex background of different degrees.

In Fig. 3a, we can see that the color of the disease leaf images is quite complex. Firstly, we convert the initial RGB images to LAB color space and then divide it into 150 superpixels by superpixel clustering. To indicate the advantage of superpixel clustering, we also segment images by only EM. The segmented lesions are shown in Fig. 3d–f.

Table 1 shows the average iterative number and average computation time to segment the four images in Fig. 3a by K -means clustering [2], fuzzy c -means clustering [10], EM algorithm [18] and superpixels + EM, where the iterative number of superpixels + EM is the total of the iterative number of all superpixels. Table 2 shows the iterative number and the computation time of superpixels + EM via different K .

In Table 2, it is found that, in experiments, for an image of the same size and threshold λ in EM algorithm, the computing time can be effectively reduced by increasing the number of superpixels K . Comparing with EM algorithm, in superpixels + EM, image is divided into K superpixels, and the iteration number increases while the computing time decreases. The reason is that EM algorithm can quickly converge in a superpixel. But, increasing K will result the over-segmentation problem. Therefore, from the overall point of view, the computation time of superpixels + EM does not increase, but reduces a lot.

In Fig. 3 and Table 1, it is seen that superpixels + EM is the best and fastest. The reason is that superpixel clustering uses the characteristic similarity between pixels to group the pixels with a small amount of superpixels instead of the pixels to express image features, which can provide the initial parameter estimation for EM, and make EM algorithm quickly converge. The segmentation speed by directly using EM is the slowest due to randomly setting the initial parameters of EM algorithm in the whole disease leaf image for image segmentation, which also leads to the phenomenon of excessive number of clusters, as shown in Fig. 3f. In Fig. 3b and f, we find that each superpixel contains a series of pixels with similar direction and similar color features, which effectively maintain a good spot boundary to some extent. Then, we segment the lesion from each superpixel by EM algorithm. Figure 3c shows the segmented lesions. In order to validate the performance of superpixels + EM, we compare it with K -means clustering and fuzzy C -means clustering algorithms.

Fig. 3 Cucumber disease leaf images and corresponding segmented lesion images. **a** Four original cucumber disease leaf images, **b** corresponding 150 superpixels by superpixel clustering, **c** segmented lesion from superpixels by superpixels + EM, **d** segmented lesion from original image by K -means clustering, **e** segmented lesion from original image by fuzzy C -means clustering, **f** segmented lesion from original image by EM

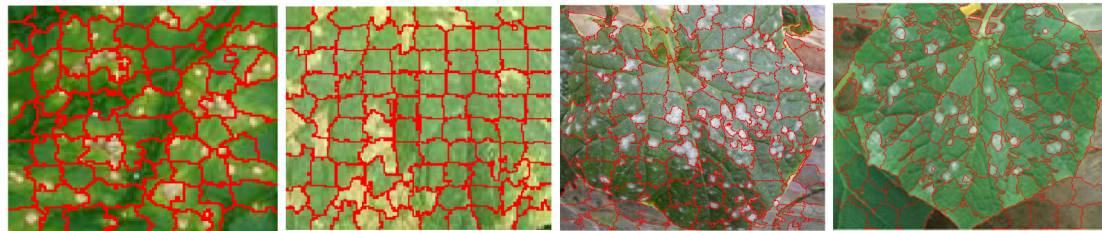
In the experiments, EM algorithm is very similar to K -means clustering. Similarly, the first step is to choose the initial parameters. But, in the two algorithms, the distances between the observation pixels and cluster centers are computed in different modules, which may lead to different results. Comparing with K -means-based disease leaf image segmentation, superpixel-based method is effective. K -means clustering adopts the squared Euclidean distance, and RGB color is chosen as the comparison parameter, while EM algorithm is not based on distances. Instead, EM algorithm computes the probability of each observation pixel belonging to each cluster based on the mixture two-Gaussian distribution, and the final goal is to find the clusters that maximize the overall probability of the data. Specially, the proposed algorithm is robust for the color image segmentation problem, because the initial parameters of EM algorithm are estimated in compact superpixel and the noise is to some extent suppressed, while the results of the segmentation methods that use a random initialization, e.g., K -means clustering and EM algorithm, may differ according to the selection of initial parameters, as demonstrated in Fig. 3d and f.

5 Conclusions and future works

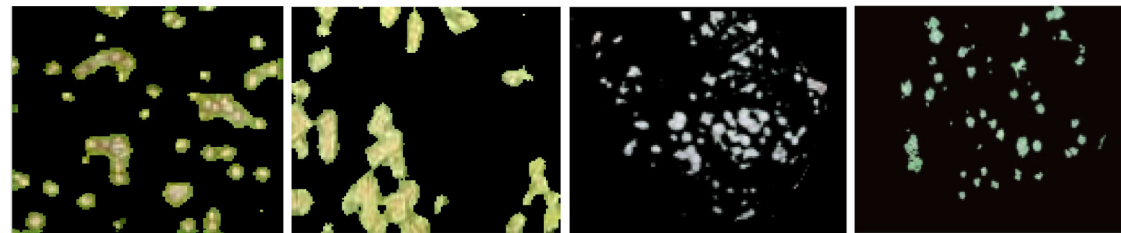
Plant disease leaf image segmentation is a crucial step in plant disease detection and recognition. In this paper, we proposed a new segmentation method of plant color disease leaf image by combining superpixel clustering and EM algorithm. Since EM algorithm is sensitive to the starting condition, we analyzed the effectiveness of applying superpixels to parameter estimation of EM algorithm. Superpixel clustering can be used to reduce the complexity of images from thousands of pixels to only a few hundreds of superpixels. In the algorithm, the color image is firstly divided into several superpixels to improve the initial estimation and possibly reduce the unlikely segmentation, and then, segmentation is carried out by EM algorithm. Experimental results show that the proposed method is appropriate for dealing with plant disease leaf image segmentation and has certain superiority in the field of plant disease detection. In the future, we will further compare the



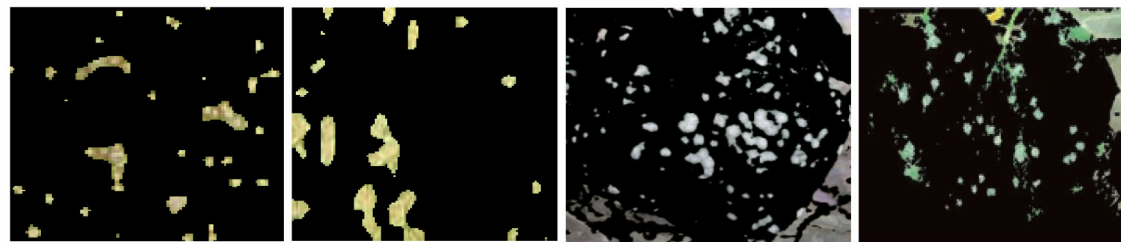
(a)



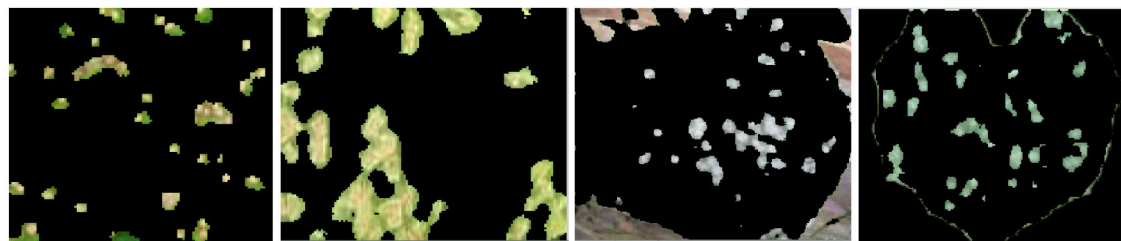
(b)



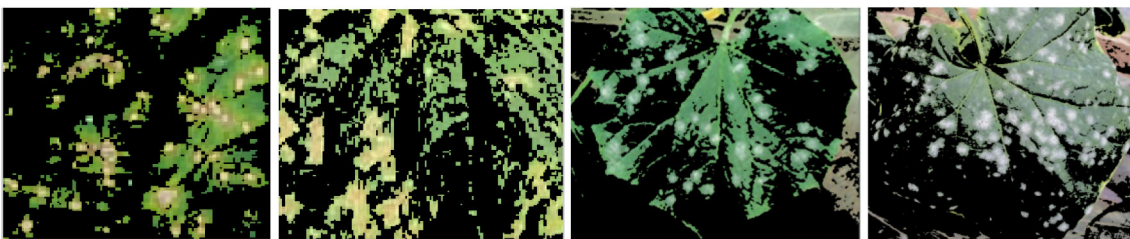
(c)



(d)



(e)



(f)

Table 1 Iterative number and the computation time of K -means clustering, EM and superpixels + EM

Method	K -means	Fuzzy C -means	EM algorithm	Superpixels + EM
Iterative number	378	475	362	659
Computation time (ms)	325	371	234	168

Table 2 Iterative number and the computation time of superpixels + EM via K

K	50	100	150	200	300	400
Iterative number	452	526	659	740	831	905
Computation time (ms)	214	198	168	157	145	114

potential benefits and limitations of the existing color image segmentation methods and explore a robust, fast and accurate disease leaf image segmentation method.

Acknowledgements This work was partially supported by China National Natural Science Foundation under Grant No. 61473237. It was also supported by the Shaanxi Natural Science Foundation Research Project under Grant No. 2016GY-141. The authors would like to thank all the editors and anonymous reviewers for their constructive advices.

Author contributions SWZ conceived the algorithm, carried out analyses, prepared the data sets, carried out experiments and wrote the manuscript; ZHY and XWW designed, performed and analyzed experiments and wrote the manuscript; all authors read and approved the final manuscript.

Compliance with ethical standards

Conflicts of interest The authors declare no conflicts of interest.

References

- Revathy R, Chennakesavan SA (2015) Threshold based approach for disease spot detection on plant leaf. *Trans Eng Sci* 3(5):72–75
- Li Guanlin M, Zhanhong HC et al (2010) Segmentation of color images of grape disease using K -means clustering algorithm. *Trans CSAE* 26(2):32–37
- Yuan Y, Li M, Liang Q et al (2011) Segmentation method for crop disease leaf images with complex background. *Trans CSAE* 27(2):208–212
- Arivazhagan S, Newlin Shebiah R, Ananthi S et al (2013) Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agric Eng Int: CIGR J* 15(1):211–217
- Gui JS, Hao L, Zhang Q et al (2015) A new method for soybean leaf disease detection based on modified salient regions. *Int J Multimed Ubiquitous Eng* 10(6):45–52
- Al-Hiary H, Bani-Ahmad S, Reyalat M et al (2011) Fast and accurate detection and classification of plant diseases. *Int J Comput Appl* (0975–8887) 17(1):31–38
- Yugang R, Jian Z, Miao L et al (2012) Segmentation method for crop disease leaf images based on watershed algorithm. *J Comput Appl* 32(3):752–755
- Kaur R, Kang SS (2015) An enhancement in classifier support vector machine to improve plant disease detection. In: *IEEE 3rd international conference on innovation and technology in education (MITE)*, 135–140
- Chaudhary P, Chaudhari AK, Cheeran AN et al (2012) Color transform based approach for disease spot detection on plant leaf. *Int J Comput Sci Telecommun* 3(6):65–70
- Hanping M, Yan cheng Z, Bo H (2008) Segmentation of crop disease leaf images using fuzzy C -means clustering algorithm. *Trans CSAE* 24(9):136–140
- Menukaewjinda A, Kumsawat P, Attakitmongkol K (2008) Grape leaf disease detection from color imagery using hybrid intelligent system. In *Proceeding of ECTI-CON 14(17)* (pp. 513–516). Krabi
- Qin F, Liu D, Sun B et al (2016) Identification of alfalfa leaf diseases using image recognition technology. *PLoS ONE* 11(12):1–26
- Baig Mohammad Md, Naga Srujana R, Jaya Naga Jyothi A et al (2016) Disease identification in plants using k -means clustering and gray scale matrices with svm classifier. *Int J Appl Sci, Eng Manage* 5(2):84–88
- Qin F, Guo J, Lang F (2015) Superpixel segmentation for polarimetric SAR imagery using local iterative clustering. *IEEE Geosci Remote Sens Lett* 12(1):13–17
- Achanta R, Shaji A, Smith K et al (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans Pattern Anal Mach Intell* 34(11):2274–2282
- Chen Q, Yang XB, Chen YH et al (2014) Image segmentation based on superpixels and saliency. *Appl Mech Mater* 701–702:312–315
- Che X, Zhang F, Zhang R (2017) Medical image segmentation based on SLIC superpixels model. *International Conference on Innovative Optical Health Science*. 1024502
- Zhaoxia Fu, Wang Liming (2012) Color image segmentation using mixture Gaussian model and EM Algorithm. *Commun Comput Inf Sci* 346:61–66
- Achanta R, Shaji A, Smith K et al (2012) SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal Machine Intell* 34(11):2274–2281
- Bai XD, Cao ZG, Wang Y et al (2013) Crop segmentation from images by morphology modeling in the CIE $L^*a^*b^*$ color space. *Comput Electron Agric* 99:21–34
- Wu CFJ (1983) On the convergence properties of the EM Algorithm. *Ann Stat* 11(1):95–103
- Einicke GA, Falco G, Malos JT (2010) EM Algorithm state matrix estimation for navigation. *IEEE Signal Process Lett* 17(5):437–440