A New Method for Soybean Leaf Disease Detection Based on Modified Salient Regions

Jiangsheng Gui, Li Hao, Qing Zhang and Xiaoan Bao

College of Information, Zhejiang SCI-Tech University, Hangzhou, China

dewgjs@126.com

Abstract

Soybean is the main food crop and an important economical crop of the world. Proper disease control measures must be undertaken to minimize losses. Techniques of machine vision and image processing were applied mostly to plant protection in recent years. Disease detection and segmentation are very important, but the diseases of soybean are complex in real environment and traditional segmentation methods cannot quickly and accurately obtain segmentation results. This research presented a new method for soybean leaf disease detection based on salient regions. This method used low-level features of luminance and color, combined with multi-scale analysis to determine saliency maps in images, and then K-means algorithm was used. The experimental results show that this method can accurately extract the disease regions from soybean disease leaf images with complex background, and it can provide an excellent foundation for extracting disease feature and identifying the diseases categories.

Keywords: artificial intelligence, K-means algorithm salient regions, image processing

1. Introduction

As we all know, in many parts of world, soybeans are the main food crop for people. But in recent years, due to some factors such as natural disasters, soil erosion and fertilizer unreasonable lead to the occurrence of crop diseases. These diseases seriously affect soybean yield and quality in some aspects. The identification disease results of traditional expert system are always influenced by human factors and this will lead to an inaccurate diagnosis. Along with the development of machine vision technology and pattern recognition, these technologies can intelligently diagnose the diseases of crop, and accurately identify the types of diseases. Image segmentation is one of the key steps, and the precision of segmentation directly affects the reliability of feature extraction and the accuracy of pattern recognition. Recently years, many methods were proposed for segmentation of crop disease image. W.Shuwen and R.Yugang proposed a watershed algorithm [1, 2] on cucumber and crop leaves disease segmentation. L. Guanlin[3] proposed a segmentation method based on k-means hard clustering algorithm for color disease image segmentation of grape. Y.Yuan proposed a method based on level set for crop disease leaves image segmentation [4]. M. Hanping [5] put forward a method based on fuzzy c-means clustering to segment the cotton disease leaves. J. Zexuan proposes a natural image segmentation method based on unsupervised fuzzy C-means (USFCM) clustering algorithm [6]. Camargo [7,8] submit a method to segment the crop disease leaves, he transform the RGB space to the HSV, I_1, I_2 and I_3 space, then extract H, I_{3a}, I_{3b} components respectively threshold segmentation for the three components. A. Meunkaewjinda [9] choose unsupervised optimal fuzzy C-means clustering (UOFCM) algorithm to segment grape leaf disease image. S. Arivazhagan proposed a color transformation structure

for the input RGB image, and then the green pixels were masked and removed using specific threshold value followed by segmentation process [10]. However the environments of segmentation presented above are controlled in many aspects. For examples, they had no complex background and only single leaf was examined. Meanwhile, they cannot achieve a desirable result by applying those methods.

Saliency detection has appeared in 90s of the last century. The most representative method is proposed by Itti *et al.* [11], who had built a computational model of saliency-based spatial attention derived from a biologically plausible architecture. They compute saliency maps for features of luminance, color, and orientation at different scales that aggregate and combine information about each location in an image and feed into a combined saliency map in a bottom-up manner. The saliency maps produced by Itti's have been used by other researchers for applications like adapting images on small devices and unsupervised object segmentation. Many scholars [12-14] in recent years proposed many improved methods to detect saliency regions based on Itti method. These methods can quickly find marked or interest objects from a large amount of image information, and effectively improved the efficiency of image processing. Therefore it has a wide range of applications in the image retrieval, image segmentation and sensitivity image recognition. Soybean disease images are made up of the disease region, healthy region and complex field background. The contrast of disease region and the other parts is larger, while there is little contrast ration between each other disease region. This work presents soybean leaf disease detection system by using color image with complex background. The innovation of this paper is that saliency regions detection was applied to divide and extract the soybean crop disease. Drawing on the idea of Itti method, we present a novel method to determine salient regions in images using low-level features of luminance and color. Then according to the K-means algorithm and combined experience threshold, salient regions can be extracted from soybean leaf disease images based on salient maps. Finally, it is necessary to correct the extraction diseases by using binary segmentation.

2. Principle of Itti Method

The main idea of Itti method was that the first input image was represented as nine levels Gaussian pyramid, wherein zero level represented the input image, from one to eight levels were formed by Gaussian filtering and sampling the input image. Then each level image was decomposed into brightness, color and direction and so on as a series of channels. A variety of visual features such as brightness, red, green, blue, yellow and direction were extracted from each level of the pyramid, then form luminance pyramid I (σ), chrominance pyramids R (σ), G (σ), B (σ), Y (σ) and direction pyramid O (θ, σ). Wherein, σ represent nine different scales, its value is from 0 to 8.So this features were expressed as nine characteristic pyramids, they contain one intensity pyramid, four color pyramids (respectively red, green, blue and yellow) and four direction pyramids (namely four different directions). By fusion calculation of all the characteristics, we could obtain saliency maps which are corresponding to the input image. Figure 1 is a block diagram of the algorithm Itti.

Since Itti method kept using neighbor interpolation algorithm, resulting in reducing the resolution of the saliency maps. Therefore, it is difficult to segment image by directly using the saliency maps of Itti approach.
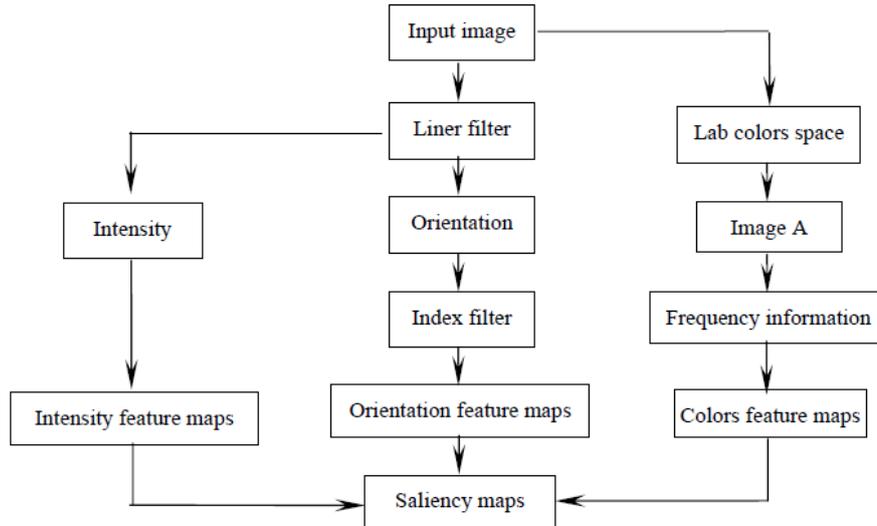
Figure1. Diagram of Itti Model

3. Soybean Leaf Disease Extraction Based on Salient Regions

3.1 Saliency Map of Soybean Disease Image

Compared with original Itti Method, we presented a new method to determine salient regions in images using low-level features of luminance and color in CIELab color space, combined with multi-scale analysis [15] by changing the size of each scale of the filter, and then generated the saliency map.

In our work, saliency is determined as the local contrast of an image region with respect to its neighborhood at various scales. This is evaluated as the distance between the average feature vectors of the pixels of an image sub-region with the average feature vector of the pixels of its neighborhood. This allows to obtain a combined feature map at a given scale by using feature vectors for each pixel, instead of combining separate saliency maps for scalar values of each feature.
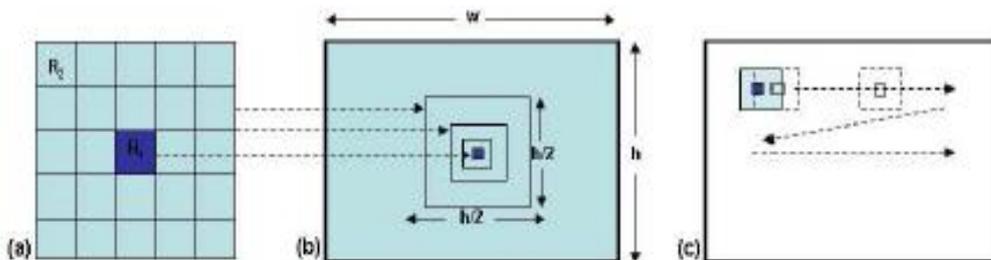

Figure2. (a) Contrast Detection Filter Showing Inner Square Region R_1 and Outer Square Region R_2. (b) The Width of R_1 Remains Constant While that of R_2 Ranges According to Equation 3 by Halving it for Each New Scale. (c) Filtering the Image at one of the Scales in a Raster Scan Fashion.

At a given scale, the contrast saliency value $c_{i,j}$ for a pixel at position (i, j) of the image is determined as the distance D, which is between the average vectors of pixel features in the inner region R1 and those in the outer region R2 (Figure 2). The Equation as follows:

$$c_{i,j} = D[(N_1 \sum_{p=1}^{N_1} V_p),(N_2 \sum_{p=1}^{N_2} V_q)] \tag{1}$$

where N_1 and N_2 are the number of pixels in R_1 and R_2 respectively, and v is the vector of feature elements corresponding to a pixel. The distance D is a Euclidean distance. Since perceptual differences in CIELab color space are approximately Euclidian, D in Equation (1) is:

$$D = \left\| V_1 - V_2 \right\| \tag{2}$$

where $V_1 = [L_1,a_1,b_1]^T$ and $V_2 = [L_2,a_2,b_2]^T$ are the average vectors for regions R_1 and R_2, respectively. Region R_1 is usually chosen to be one pixel, and if the image is noisy then R_1 can be a small region of N *N pixels. For an image of width w pixels and height h pixels, the width of region R_2, namely w_B is varied as:

$$\frac{w}{8} \leq w_B \leq \frac{w}{2} \tag{3}$$

assuming w to be smaller than h (else we choose h to decide the dimensions of R_2).So for each image, filtering was performed at three different scales (according to Eq. 3) and the final saliency map is determined as a sum of saliency values across the scales S:

$$m_{i,j} = \sum_S C_{i,j} \ (i \in [1,w], j \in [1,h]) \tag{4}$$

where $m_{i,j}$ is an element of the combined saliency map M obtained by point-wise summation of saliency values across the scales.

3.2 Diseases Segmentation Using Saliency Maps

The following describes how to extract a saliency area based on the saliency maps. The image was over-segmented by using a simple K-means algorithm. The K seeds for the K-means segmentation were automatically determined by using the hill-climbing algorithm [16] in the three-dimensional CIELab histogram of the image. The hill-climbing algorithm can be seen as a search window being run across the space of the d-dimensional histogram to find the largest bin within that window. Since the CIELab feature space is three-dimensional, each bin in the color histogram has $3^d - 1 = 26$ neighbors where d is the number of dimensions of the feature space. The number of peaks obtained indicates the value of K, and the values of these bins form the initial seeds.

Since K-means algorithm clusters pixels were in the CIELab feature space, an 8-neighbor connected-components algorithm was run to connect pixels of each cluster spatially. Once the segmented regions r_k for k = 1; 2; …K was found, the average saliency value V per segmented region was calculated by adding up values in the final saliency map M corresponding to pixels in the segmented image:

$$P_k = \frac{1}{|r_k|} \sum_{i,j \in r_k} m_{i,j} \tag{5}$$

where $|r_k|$ is the size of the segmented region in pixels. A simple threshold based method can be used, wherein the segments those average saliency value are greater than a

certain threshold T were retained, and while the rest are discarded. These results in an output contain only those segments that constitute the salient object.

3.3 Correction for the Diseases Extraction

Compared to Itti method and some methods on the basis of its improvement, above method can more quickly and accurately extract significant area. But it still exist a common problem, when the contrast of the salient regions and the surrounding area is not obvious, it will lead to misjudgment, resulting in a non-salient region was be preserved. Crop disease, as special objects, are required the ability to be more completely and accurately segmentation of disease region, and provide an excellent foundation for extracting characteristic parameters, identifying and diagnosing the diseased categories. Since the contrast of leaf vein parts and other parts is little, disease areas were extracted will contain some irrelevant parts, thereby it will affect the accuracy of disease segmentation. Inaccurate segmentation will affect the precise of the subsequent parameters. So it is needed to correct the disease area after extracted. The color of the disease area is very complex, and it is difficult to describe by any color in the RGB color space. LAB color space is a color mode of perceptual uniform, which image contains all of the color information in the AB two-dimensional space. This research combined two-color space to correct salient regional. Firstly, threshold segmentation was applied to R component of RGB color space, then using morphology algorithm to fill hole of salient regions and remove the small size parts of some non-disease area. Through above processing, the picture only contains some lager healthy blade departments. Finally it is need to convert the initial RGB components towards to LAB color space, a complete disease region can be achieved by applying binary segmentation to a component of LAB color space.

4. Experiments and Results

Experiments were conducted in a mixture platform of Matlab7.0 and VC++. All the pictures of the experiments are from in Hefei academy. For segmentation, a window 3*3*3 was used for the hill-climbing search on a 16*16*16 bin CIELab histogram. The average saliency threshold T which was used for selecting segments was set at 25 (about 10% of the maximum possible average saliency in the normalized final saliency map).

Figure 3 is 6 soybean leaf images with complex background of different degree, and the result images conducted by methods described in this research. It can be seen that the method based on salient region can be quickly and accurately enough to locate the area of diseases in complex farmland background. Figure 3 (c) is the result of disease segmentation by means of correcting the extracted disease area. Compared these two images, it is easy to find that veins and textures in the corrected figure were took out. Thus the accuracy of segmentation is high, and ensures the reliability of disease characteristic parameters extraction and accuracy of pattern recognition in the succeeding work.
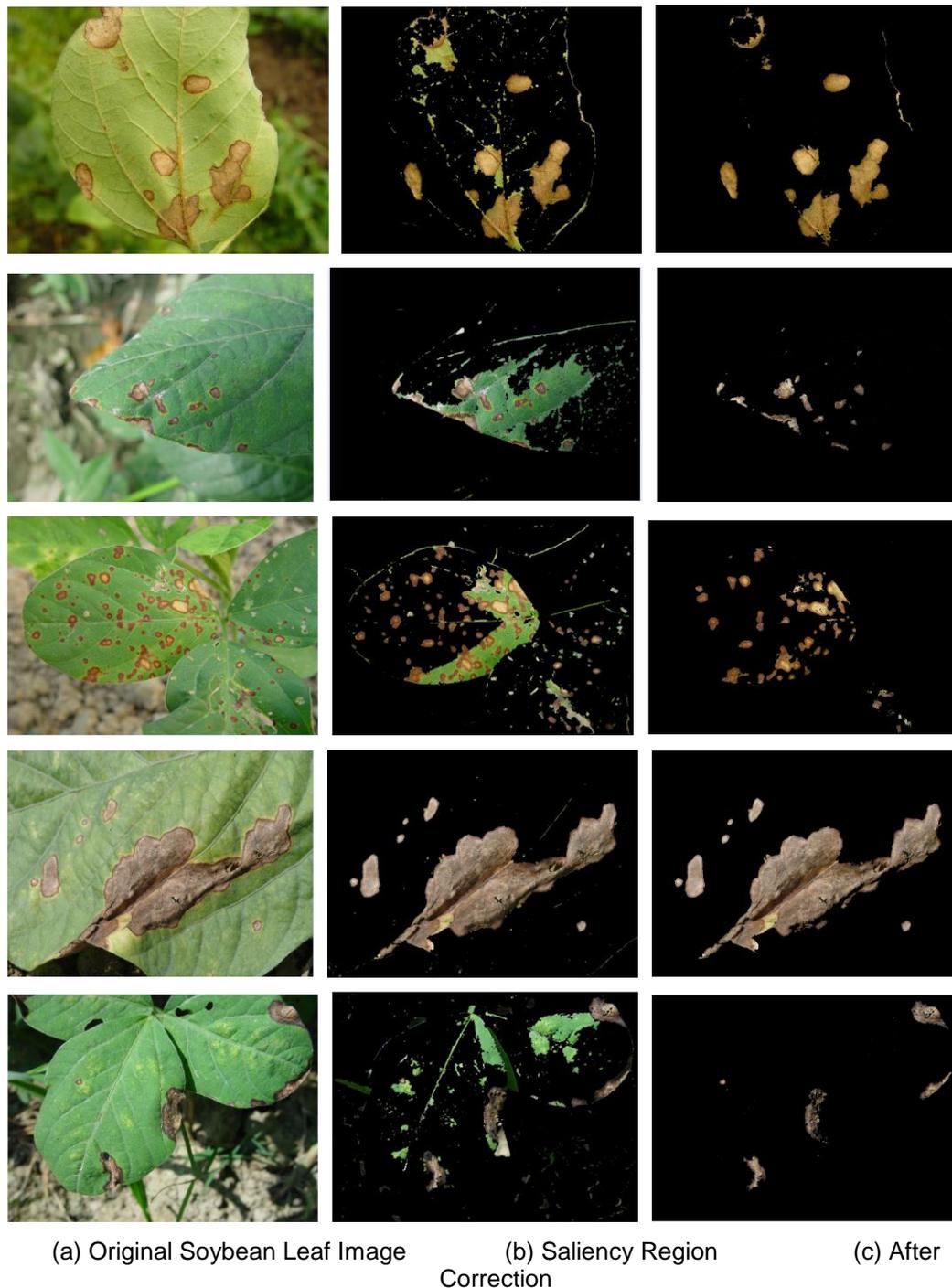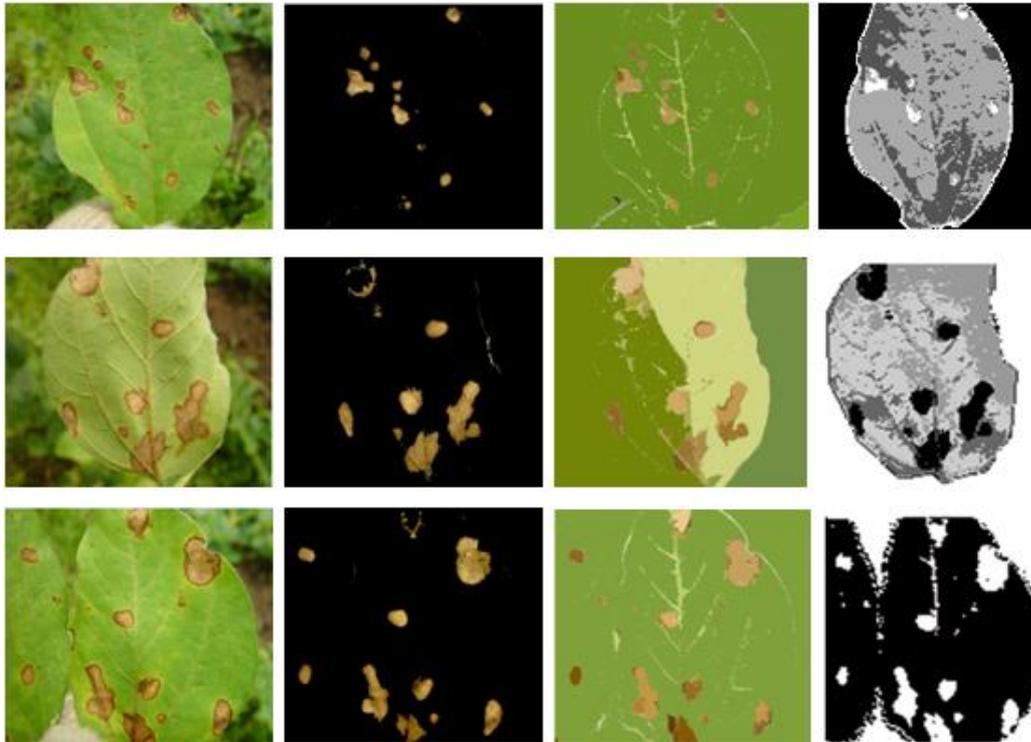
(a) Original Soybean Leaf Image   (b) Saliency Region Correction   (c) After Correction
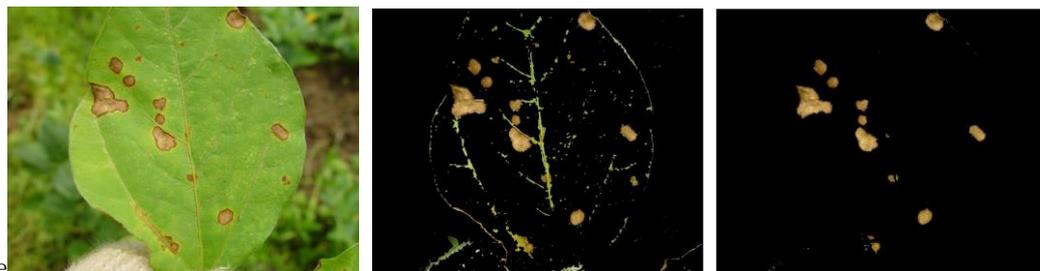
Figure 3. Contrast Disease Results Before and After Correction

In order to highlight advantages of the described method in this research, Figure 4 shows the results of salient region segmentation of our method were compared to Mean Shift algorithm and unsupervised optimal fuzzy C-means clustering for the same soybean disease image. It can be seen that directly using Mean Shift algorithm will lead to the phenomenon of excessive number of categories, and unsupervised fuzzy optimal c-means clustering (UOFCM) cannot effectively overcome the effect on the segmentation results of the texture. Experimental results show that the proposed method can achieve a more favorable segmentation effects. As explained further in this research, this method is appropriate for dealing with segmentation crop disease images. In addition, it has also

illustrated that the method based on salient region extraction has certain superiority in the field of crop disease image segmentation.

(a)Original soybean leaf image (b) the proposed method (c) Mean shift (d) UOFCM

Figure 4. Segmentation Results Comparison by 3 Methods

5. Conclusions

We presented a novel method to extract crop disease region based on saliency detection, which is easy to implement and noise tolerant. In our experiments, we directly extracted the disease and removed the step of segmentation background, so that the complexity and difficulty of experiments were reduced. The experimental results show that the proposed method gives efficient soybean leaf disease extraction with complex background, and it produced better results than the other methods such as Mean shift and UOFCM.

Acknowledgment

References

[1] W. Shuwen, Z. Changl,"Study on identification of cucumber leaf disease based on image processing", Journal of Northeast Agricultural University, vol.43 no.5, **(2012)**, pp.69-73.
[2]   R. Yugang, Z. Jian, L. Miao and Y. Yuan, "Segmentation method for crop disease leaf images based on watershed algorithm", Journal of Computer Applications, vol. 32 no.3, **(2012)**, pp.752-755.
[3]   Li Guanlin, M. Zhanhong, H. Chong, C. yongwei and W. Haiguang, "Segmentation of color images of grape disease using K_means clustering algorithm", Transactions of the CSAE, vol.26 no.2, **(2010)** , pp.32-37.

[4]   Y. Yuan, Li Miao, Q. Liang, H. Xiuzhen and Z. Wei," Segmentation method for crop disease leaf images with complex background", Transactions of the CSAE, vol.27 no.2, **(2011)**,pp. 208－212.

[5]   M. Hanping, Z. Yan cheng, H. Bo, "Segmentation of crop disease leaf images using fuzzy C-means clustering algorithm", Transactions of the CSAE, vol.24 no.9, **(2008)**, pp.136－140.

[6]   J. Zexuan, P. Yu, C. Qiang, S. Quansen and X. Deshen, "Natural image segmentation algorithm with unsupervised FCM", Journal of Image and Graphics, vol.16 no.5, **(2011)**, pp.773-783.

[7]   A. Camargo, J. S. Smith, "Image pattern classification for the identification of disease causing agents in plants", Computers and electronics in Agriculture, vol.66 no.2, **(2009)**, pp. 121-125.

[8]   A.Camargo, J.S.Smith, "An image processing based algorithm to automatically identify plant disease visual symptoms", Biosystems engineering, vol 102.no.1, **(2009)**, pp.9-21.

[9]   A. Menukaewjinda, P. Kumsawat, K. Attakitmongcol, "Grape leaf disease detection from color imagery using hybrid intelligent system", Proceeding of ECT1-CON, Krabi, Thailand, **(2008)**, May,14-17, pp.513-516.

[10]   S. Arivazhagan, R. Newlin Shebiah, S. Ananthi and S. Vishnu Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features", Agricultural Engineering International: CIGR Journal, vol.15 no. 1, **(2013)**, pp. 211-217.

[11]   L. Itti, C. Koch, E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.22 no.11, (**1998**), pp. 1254-1 260.

[12]   Y. Ming, W.Qian, G. Ying-chun, "A Method of Extraction of Salient Regions in Image", Opto-Electronic Engineering, vol.39 no.8, (**2012**), pp.18-25.

[13]   R.Achanta, F.Estrada, P.Wils and S.Ä usstrunk, "Salient Region Detection and Segmentation", International Conference on Computer Vision Systems (ICVS '08), Springer Lecture Notes in Computer Science, Santorini,Greece, (**2008**), May 12-15, pp.66-75.

[14]   E. Rahtu, J. Kannala, M.Salo and J. Heikkila, "Segmenting Salient Objects from Images and Videos", 11th European Conference on Computer Vision, Crete, Greece, (**2010**), September 5-11, pp.366-379.

[15]   Z. Qiaorong, G. Guochang, L. Haibo and X. Huimin, "Salient region detection using multi-scale analysis in the frequency domain", Journal of Harbin Engineering University, vol.36 no.3, (**2010**), pp. 361:365.

[16]   T. Ohashi, Z. Aghbari, and A. Makinouchi, "Hill-climbing algorithm for efficient color-based image segmentation". Proceeding of IASTED International Conference On Signal Processing, Pattern Recognition, and Applications (SPPRA), Rhodes, Greece, (**2003**), Jun.30-July.

Authors
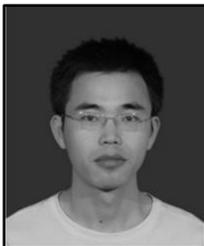
Jiangsheng Gui, He received his Bachelor in Mathematics (2000) and PhD in Biosystem engineering (2007) from Zhejiang University China. Now he is full associate professor of informatics at Informatics Department, Zhejiang Sci-Tech University. His current research interests include different aspects of Image Processing and Computer Vision.