International Research Journal of Management, IT & Social Sciences
Available online at https://sloap.org/journals/index.php/irjmis/
Vol. 5 No. 1, January 2018, pages: 86~94
ISSN: 2395-7492
https://sloap.org/journals/index.php/irjmis/article/view/897

Automatic Rice Leaf Diseases Detection Using SVM

Hema Rajini [a]

Article history:

Received: 10 July 2017
Revised: 12 January 2018
Approved: 15 January 2018
Published: 27 January 2018

Keywords:

expectation maximization;
pyramid of histograms of orientation gradients;
support vector machine;

Abstract

Rice leaf diseases can be detected and recognized automatically. This proposed method, combines super pixels, expectation maximization algorithm, and pyramid of histograms of orientation gradients, to recognize rice diseases. In the proposed method first, pre-processing is performed. Second simple linear iterative clustering is used to divide a diseased leaf image into a number of compact regions, which can dramatically accelerate the convergence speed of the expectation maximization algorithm that is adopted to segment the diseased leaf regions and obtain the lesion image. Third, the pyramid of histograms of orientation gradients features are extracted from the segmented lesion image. In the fourth stage extracted features are reduced using principal component analysis Finally, support vector machine is used to classify and recognize different rice diseases. A database of rice diseased leaf images, is taken to conduct the experiment and the results show that the proposed method is effective and feasible for recognizing rice diseases.

Author correspondence:
Hema Rajini,
Department of Computer Science and Engineering, Alagappa Chettiar Government College of Engineering and Technology, Karaikudi – 630003, Tamilnadu, India.
Email address: auhemasmith@yahoo.co.in

1. Introduction

Rice (Oryza Sativa) is one of the important crops in India. The majority of the land area is under the cultivation of the rice crop and is one of the main sources of economic development. Every year farmers face a loss of yield and financial losses due to the pests and the diseases on the rice plants. Bacteria, fungi, and viruses are the main reasons of the rice diseases. There are many rice diseases, out of which four diseases have been used in this research work namely Rice Bacterial Blight (RBB), Rice Blast (RB), Rice Brown Spot (RBS) and Rice Sheath Rot (RSR). These diseases have some similar symptoms which confuse the human vision while detecting them. Early symptoms of some diseases cannot be detected by the farmers. Farmers require continuous monitoring by experts in order to diagnose diseases at an early stage which are very expensive and time-consuming. This creates a need for image processing techniques to detect diseases automatically.

[a] Alagappa Chettiar Government College of Engineering and Technology, Karaikudi – 630003, Tamilnadu, India

Image processing techniques help not only to diagnose the diseases accurately but also to classify diseases. Early and accurate identification of the diseases is also possible. This will increase the productivity and quality of the rice yield. With image processing techniques, human efforts are reduced in great extent. Image processing techniques are widely used for different agricultural applications ranging from identification of the leaflet of the plant to categorization of different diseases. Rice disease recognition based on diseased leaf images has become an increasingly important and challenging research topic in the community.

There are many computer-aided detection systems for leaf diseases in the literature, most of them are used to segment and classify plant leaf diseases. Kiran et al. reviews and summarizes image processing techniques for several plant species that have been used for recognizing plant diseases[1]. The major techniques for detection of plant diseases are: backpropagation neural network, support vector machine, K-means clustering, and probabilistic neural networks. These techniques are used to analyses the healthy and diseased plants leaves. Along with the development of image processing, computer vision and pattern recognition, many rice disease recognition methods have been summarised.

Detection of unhealthy region and classification using texture features have been proposed[2]. Their algorithm has been tested on ten species of plants namely banana, beans, jackfruit, lemon, mango, potato, tomato and sapota. 94.74% accuracy has been achieved by Support vector machine (SVM) classifier. Tried another ANN, i.e. back propagation neural network (BPNN) for efficient grape leaf colour extraction with complex background[3]. They also explore modified self-organizing feature map (MSOFM) and genetic algorithm (GA) and found that these techniques provide automatic adjustment in parameters for grape leaf disease colour extraction. SVM has been also found to be very promising to achieve efficient classification of leaf diseases. 21 colour, 4 shape and 25 texture features has been extracted by Wang et al. and principal component analysis (PCA) has been performed for reducing dimensions in feature data processing, then back-propagation (BP) networks, radial basis function (RBF) neural networks, generalized regression networks (GRNNs) and probabilistic neural networks (PNNs) has been used as the classifiers to identify diseases[4].

Developed an algorithm for image segmentation technique which is used for automatic detection and classification of plant leaf diseases[5]. It also covers survey on different diseases classification techniques that can be used for plant leaf disease detection. Image segmentation, which is an important aspect for disease detection in plant leaf disease, is done by using genetic algorithm. Proposed a method to detect rice diseases of downy mildew, powdery mildew and anthracnose using leaf image processing and recognition technologies[6]. They extracted features, such as colour, shape and texture using gray level co-occurrence matrix, and identified the diseases by the nearest neighbour classifier. Segmented the rice disease images and obtained the lesions, and extracted the features of colour and texture from the representative lesion with maximum area[7].

2. Research Methods

Developing an efficient detection and classification method helps user to know the affected area and type of leaf diseases in an appropriate time. The proposed method has five stages, namely pre-processing, segmentation, feature extraction, feature reduction and classification. The proposed technique for automatic leaf disease classification is illustrated in Figure 1. A rice disease recognition method developed by fusion of super pixel, EM algorithm and pyramid of histograms of orientation gradients. The whole colour 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 algorithm, and then, the lesion pixels are quickly and accurately segmented from each superpixel by EM algorithm.

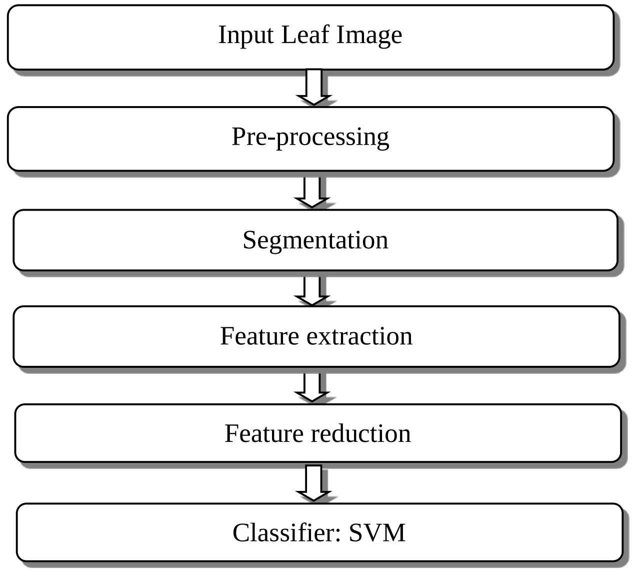

Figure 1. Methodology of the Proposed Technique

Image Preprocessing

First, select the rice plant which is affected by the disease and then collect the leaf of the plant and take a snapshot of leaf and load the leaf image into the system. In this stage to remove noise median filter is used. Figure 2 shows the input leaf image.

Figure 2. Diseased leaf image

SLIC Segmentation

Simple Linear Iterative Clustering (SLIC) is widely applied to superpixel clustering due to its simplicity and practicality [8]. 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 centre for each pixel and then iteratively assigns pixels to the closest cluster centre and updates the locations of the cluster centres.

A super pixel is an image patch which is better aligned with intensity edges than a rectangular patch. Super pixels can be extracted with any segmentation algorithm; however, most of them produce highly irregular super pixels, with widely varying sizes and shapes. Super pixel is a group of connected pixels with similar colours or gray levels. Super pixel segmentation is dividing an image into hundreds of non-overlapping super pixels. SLIC is a simple and efficient method to decompose an image in visually homogeneous regions. It is based on a spatially localized version of k-means clustering. Similar to mean shift or quick shift, each pixel is associated to a feature vector.

Expectation-Maximization Algorithm

Expectation Maximization (EM) algorithm consists of E-step and M-step[9]. 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 Estep. 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 modelled by a probabilistic distribution. In general, for each superpixel, the region is small, and the pixel colour is similar and compact, and the colour 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.

Pyramid Histogram of Oriented Gradient Features (PHOG)

Pyramid Histogram of Oriented Gradient Features (PHOG) feature is a spatial shape descriptor applied to image classification recently[10]. It represents the spatial distribution of edges and is formulated as a vector representation. This descriptor is mainly inspired by two sources: (1) the use of the pyramid representation, and (2) the Histogram of Orientation Gradients (HOG). The PHOG features are extracted from the leaf.

The details of extracting PHOG features are as follows.

Step 1: Extracting edge contours. Giving a sample image, edge contours of the image are extracted first for further processing. The edge contours were extracted using the Canny edge detector.

Step 2: The image is divided into cells at several pyramid level. The grid at level L has 2^L cells along each dimension.

Step 3: The HOG for each grid at each pyramid resolution level was computed. The Local level intrinsic shape is represented by a grayscale histogram of internal edge orientations within an available image subregion quantized into K bins. Edge contours were located in step 1, and the orientation gradients were computed at edge contours in the original image. The orientation gradients were computed using 3x3 Sobel mask without Gaussian smoothing.

Step 4: The final PHOG descriptor for an image is a concatenation of all the HOG vectors at each pyramid resolution. The concatenation of all the HOG vectors introduces the spatial information of the image. Each HOG is normalized to sum to unity taking into account all the pyramid levels.

Pyramid histogram of oriented gradient features (PHOG) is a part-based crescent like structure, whose simplicity combined with an original learning strategy leads to a fast and high accuracy detect results. Histogram of Oriented Gradient (HOG) can capture edge or gradient structure that is very characteristic of local shape, and gets better invariance to local geometric and photometric transformations by using gradient and histogram normalization This not only simplifies the classification procedure, but also removes the influence of botany conception that may change all the time.

Feature reduction using Principal Component Analysis

Principal Component Analysis (PCA) is one of the statistical techniques frequently used in signal processing to the data dimension reduction or to the data correlation[11]. The PCA and independent component analysis (ICA) are two well-known tools for transforming the existing input features into a new lower dimensional feature space. In PCA, it is transformed using the largest eigenvectors of the correlation matrix. In the ICA, the original input space is transformed into an independent feature space with a dimension that is independent of the other dimensions. PCA is the most widely used subspace projection technique. These methods provide suboptimal solution with a low computational cost and complexity. The PHOG coefficients are the set of data in which PCA finds the linear lower-dimensional representation of the data so that the variance of the reconstructed data is preserved. Using a system of feature reduction based on PCA limits, the feature vectors to the component selected by the PCA, which leads to an efficient classification algorithm. So, the main idea behind using PCA in our approach is to reduce the dimensionality of the PHOG features, which results in a more efficient and accurate classifier. The size of the input matrix is reduced from (1,746) to (7).

ISSN: 2395-7492

Classification using SVM

Support vector machine (SVM) is a powerful supervised classifier and accurate learning technique. From the statistical theory it was derived and developed by Vapnick in 1982. It yields successful classification results in various application domains, e.g. medical diagnosis. SVM is based on the structural risk minimization principle from the statistical learning theory[12]. The kernel controls the empirical risk and classification capacity in order to maximize the margin between the classes and minimize the true costs. SVM searches an optimal separating hyper-plane between members and non-members of a given class in a higher dimensional feature space. The inputs to the SVM algorithm are the features reduced by PCA. In this method, four classes of rice leaves are used.

3. Results and Analysis

The rice diseased leaf images are complex, which are composed of diseased leaf regions, healthy leaf regions and complex background. Moreover, the symptoms of a diseased leaf are different while the plant diseases are developing; the colour, size, shape and spot distribution of a diseased leaf symptom are not always the same, which makes the plant disease recognition.

Lab colour space is a 3-axis colour system with dimension L for lightness and a and b for the colour dimensions. Working with the Lab colour space includes all of colours in the spectrum, as well as colours outside of human perception. The Lab colour space is the most exact means of representing colour and is device independent. This accuracy and portability make it suitable in a number of different industries such as printing, automotive, textiles, and plastics. Lab colour is usually converted to less accurate colour spaces, such as RGB and CYMK, because computer monitors and printers use either three or four colours to represent images. Figure 3 shows Lab colour space.

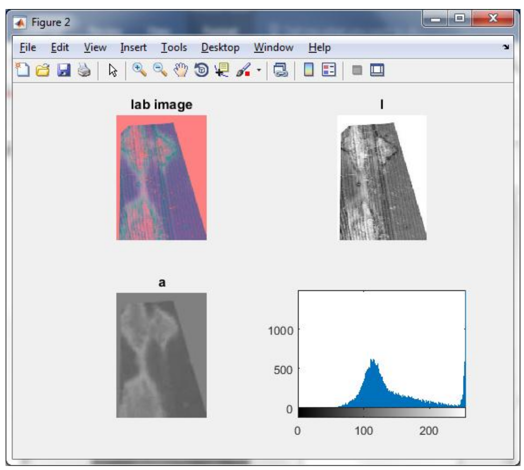

Figure 3. Lab Colour Space

A histogram is an accurate graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable and was first introduced by Karl Pearson. It is a kind of bar graph. The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often (but are not required to be) of equal size. Figure 4, figure 5 and figure 6 shows histogram for L, A and B.

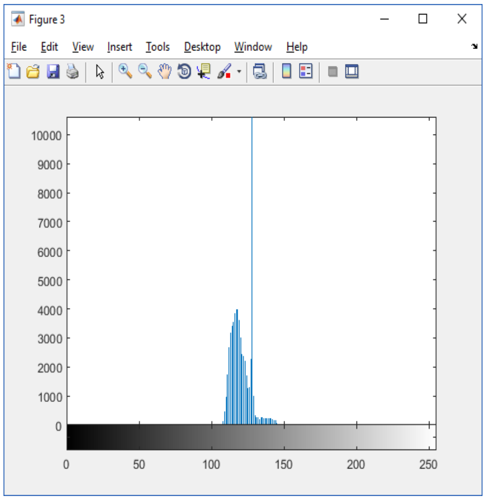

Figure 4. Histogram for L

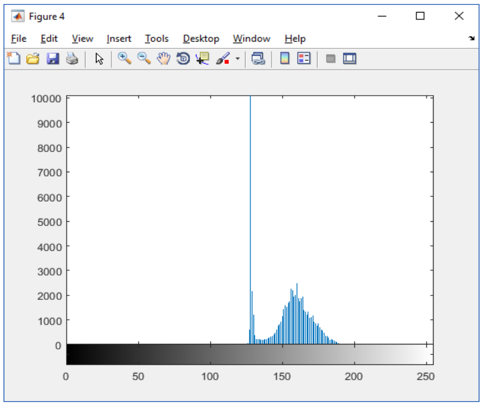

Figure 5. Histogram for A

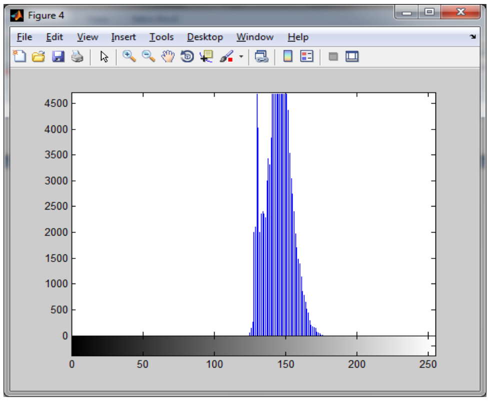

Figure 6. Histogram for B

A super pixel can be defined as a group of pixels which have similar characteristics. It is generally colour based segmentation. Super pixels can be very helpful for image segmentation. There are many algorithms available to segment super pixels but the one that used is state of the art with a low computational overhead. Simple linear iterative clustering is the algorithm to segment super pixels which doesn't require much computational power. Figure 7 presents the result of super pixel segmentation.

The method firstly convert image from RGB colour space to HSV Colour Space; Secondly, make use of a model of mixture K Gaussians, the EM formula is used to estimate the parameters of the GMM, which the desired number of partitions and fits the image histogram. The EM algorithm tends to get stuck less than K-means algorithm. The idea is to assign data points partially to different clusters instead of assigning to only one cluster. To do this partial assignment, model each cluster using a probabilistic distribution.

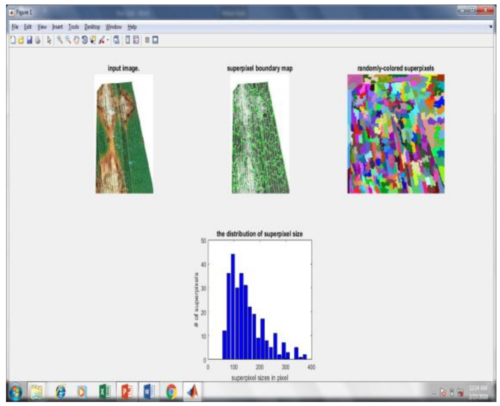

Figure 7. Super Pixel Segmentation

This algorithm uniformly utilizes fixed thresholds from diseased leaf image with gray level differences between the healthy pixels and spot pixels. In reality, however, the healthy part and spot part in a diseased leaf image are fuzzy, and the colour of healthy part and spot part is often uneven and unclear. The overlapping clearly makes it difficult to determine the uncertain lesion pixels from diseased leaf image by a fixed threshold, criteria or its grayscale histograms. First SLIC is applied to divide the original diseased leaf into several super pixels; Second EM is used to segment each super pixel and obtain the lesion image. Figure 8 shows the leaf image after segmentation.

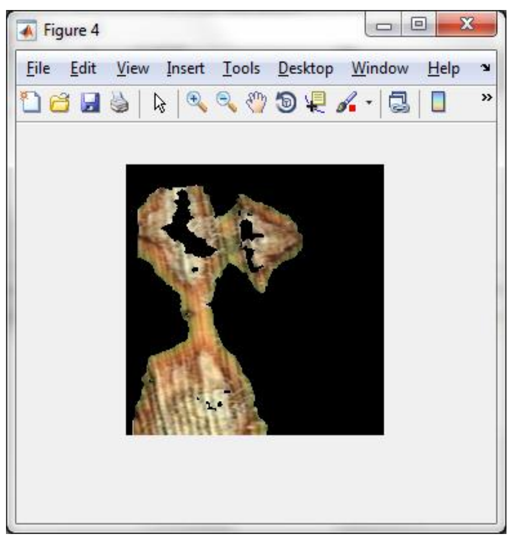

Figure 8. Segmentation of EM Algorithm

Feature extraction by PHOG can capture edge or gradient structure that is very characteristic of local shape, and gets better invariance to local geometric and photometric transformations by using gradient and histogram normalization. This not only simplifies the classification procedure, but also removes the influence of botany conception that may change all the time. The PHOG features extracted is 1746. Then these features are reduced using PCA to 7. Finally, SVM is used to classify the images. Thus, this classifier achieves 95% of classification accuracy.

4. Conclusion

The rice disease recognition method is developed based on fusing super pixels, EM algorithm and PHOG. Super pixels are employed in to help EM clustering algorithm to initialize the parameters by greatly reducing the redundancy and complexity of the diseased leaf image processing. After the parameter estimation by EM, the lesion image is obtained by the conventional maximum likelihood method. The robust PHOG features are extracted. Extracted features are reduced by PCA. SVM is applied to disease recognition. The proposed method is validated on a rice diseased leaf database. The experimental results show that the proposed method is efficient and feasible for the rice disease recognition. Finally, support vector machine is used to classify the leaf images. It is the efficient technique and it achieves 95% of classification accuracy.

Acknowledgements

I am grateful to the two anonymous reviewers for their valuable comments on the earlier version of this paper.

References

1. Kiran, R & Gavhale, U. G. (2014). An overview of the research on plant leaves disease detection using image processing techniques. IOSR Journal of Computer Engineering, 16 (1), 10–16.
2. Arivazhagan, S., Newlin Shebiah, R., Ananthi, S. & Vishnu Varthini, S. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Commission Internationale du Genie Rural Journal, 15(1), 211-217.
3. Menukaewjinda, A., Kumsawat, P., Attakitmongcol, K. & Srikaew, A. (2008). Grape leaf disease detection from colour imagery using hybrid intelligent system. Proceedings of electrical Engineering/electronics, Computer, Telecommunications and Information technology, Krabi, Thailand, 1, 513-516.
4. Wang, H., Li, G., Ma, Z. & Li, X. (2012). Image Recognition of Plant Diseases Based on Principal Component Analysis and Neural Networks. 8th International Conference on Natural Computation, Chongqing, China, 246-251.
5. Vijai Singh & Misra, A.K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture, 4(1), 41-49.,
6. Pixia, D., Xiang, D. & Wang. (2013). Recognition of greenhouse rice disease based on image processing technology. Open Journal of Applied Sciences, 3, 27–31.
7. Guo, P., Liu, T. & Li, N. (2014). Design of automatic recognition of rice disease image. Information Technology Journal, 13(13), 2129–2136.
8. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Süsstrunk, S. (2012). SLIC Superpixels compared to state-of the-art superpixel methods. IEEE Transactions on Pattern Analysis & Machine Intelligence, 34(11):2274–2281
9. Jeff Wu, C. F. (1983). On the convergence properties of the EM Algorithm. Annals of Statistics, 11(1):95–103
10. Sonali, N. (2017). Leaf Recognition Based on PHOG and COALBP Features. International Journal for Research in Applied Science & Engineering Technology, 5(6), 2751-2755.
11. Jolliffe, I. T. (1986). Principal component analysis. Springer, Berlin, 487.
12. Cristianini, N., & Shawe Taylor, J., (2000). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, First Edition, Cambridge University Press, England.