Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features

S. Arivazhagan, R. Newlin Shebiah*, S. Ananthi, S. Vishnu Varthini

(*Department of Electronics and Communication Engineering*, *Mepco Schlenk Engineering College*, *Sivakasi Tamilnadu*, *626 005*, *India*)

Abstract: Plant diseases have turned into a dilemma as it can cause significant reduction in both quality and quantity of agricultural products. Automatic detection of plant diseases is an essential research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases as soon as they appear on plant leaves. The proposed system is a software solution for automatic detection and classification of plant leaf diseases. The developed processing scheme consists of four main steps, first a color transformation structure for the input RGB image is created, then the green pixels are masked and removed using specific threshold value followed by segmentation process, the texture statistics are computed for the useful segments, finally the extracted features are passed through the classifier. The proposed algorithm's efficiency can successfully detect and classify the examined diseases with an accuracy of 94%. Experimental results on a database of about 500 plant leaves confirm the robustness of the proposed approach.

Keywords: HSI, color co-occurrence matrix, texture, SVM, plant leaf diseases

Citation: S.Arivazhagan, R. Newlin Shebiah, S.Ananthi, S.Vishnu Varthini. 2013. Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int: CIGR Journal, 15(1): 211−217.

1 Introduction

Images form important data and information in biological sciences. Digital image processing and image analysis technology based on the advances in microelectronics and computers has many applications in biology and it circumvents the problems that are associated with traditional photography. This new tool helps to improve the images from microscopic to telescopic range and also offers a scope for their analysis. It, therefore, has many applications in biology (Rastogi and Chadda, 1989).

. Plant diseases cause periodic outbreak of diseases which leads to large scale death and famine. It is estimated that the outbreak of helminthosporiose of rice in north eastern India in 1943 caused a heavy loss of food grains and death of a million people. Since the effects of plant diseases were devastating, some of the crop cultivation has been abandoned. It is estimated that 2007 plant disease losses in Georgia (USA) is approximately $653.06 million (Jean, 2009). In India no estimation has been made but it is more than USA because the preventive steps taken to protect our crops are not even one-tenth of that in USA.

The naked eye observation of experts is the main approach adopted in practice for detection and identification of plant diseases. But, this requires continuous monitoring of experts which might be prohibitively expensive in large farms. Further, in some developing countries, farmers may have to go long distances to contact experts, this makes consulting experts too expensive and time consuming (Al-Hiary et al., 2011) and moreover farmers are unaware of non-native diseases.

Automatic detection of plant diseases is an important research topic as it may prove benefits in monitoring large fields of crops, and thus automatically detect the diseases from the symptoms that appear on the plant leaves. This enables machine vision that is to provide image based

Received date: 2012-11-23 **Accepted date:** 2013-02-21

* **Corresponding author:** **R. Newlin Shebiah**, Email: newlinshebiah@yahoo.co.in.

automatic inspection, process control and robot guidance. Comparatively, visual identification is labor intensive, less accurate and can be done only in small areas.

Kim et al. (2009) have classified the grape fruit peel diseases using color texture features analysis. The texture features are calculated from the Spatial Gray-level Dependence Matrices (SGDM) and the classification is done using squared distance technique. Grape fruit peel might be infected by several diseases like canker, copper burn, greasy spot, melanose and wind scar (Kim et al., 2009).

Helly et al. (2003) developed a new method in which Hue Saturation Intensity (HIS) - transformation is applied to the input image, then it is segmented using Fuzzy C-mean algorithm. Feature extraction stage deals with the color, size and shape of the spot and finally classification is done using neural networks (Helly et al., 2003). Real time specific weed discrimination technique using multilevel wavelet decomposition was proposed by Siddiqil et al. (2009). In this histogram equalization is used for preprocessing. Features are extracted from wavelet decomposition and finally classified by Euclidean distance method (Siddiqil et.al, 2009)

Al-Bashish et al. (2011) developed a fast and accurate method in which the leaf diseases are detected and classified using k-means based segmentation and neural networks based classification. Automatic classification of leaf diseases is done based on high resolution multispectral and stereo images (Bauer et al., 2011). Sugar beet leaves are used in this approach.

Segmentation is the process that is carried out to extract the diseased region and the plant diseases are graded by calculating the quotient of disease spot and leaf areas. An optimal threshold value for segmentation can be obtained using weighted Parzen-window (Jun and Wang, 2008). This reduces the computational burden and storage requirements without degrading the final segmentation results.

In this paper, detection and classification of leaf diseases has been proposed, this method is based on masking and removing of green pixels, applying a specific threshold to extract the infected region and computing the texture statistics to evaluate the diseases. Plant diseases may be broadly classified into three types. They are bacterial, fungal and viral diseases.

2 Proposed methodology

First, the images of various leaves are acquired using a digital camera. Then image-processing techniques are applied to the acquired images to extract useful features that are necessary for further analysis. After that, several analytical techniques are used to classify the images according to the specific problem at hand. Figure 1 depicts the basic procedure of the proposed vision-based detection algorithm in this paper.

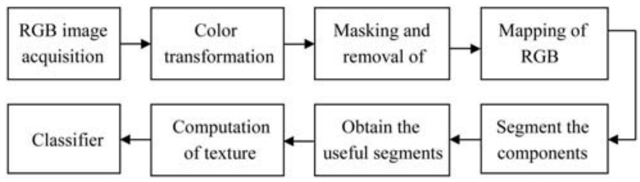


Figure 1 Block diagram of proposed approach

In the initial step, the RGB images of all the leaf samples were picked up.

The step-by-step procedure of the proposed system:

1) RGB image acquisition;

2) Convert the input image from RGB to HSI format;

3) Masking the green-pixels;

4) Removal of masked green pixels;

5) Segment the components;

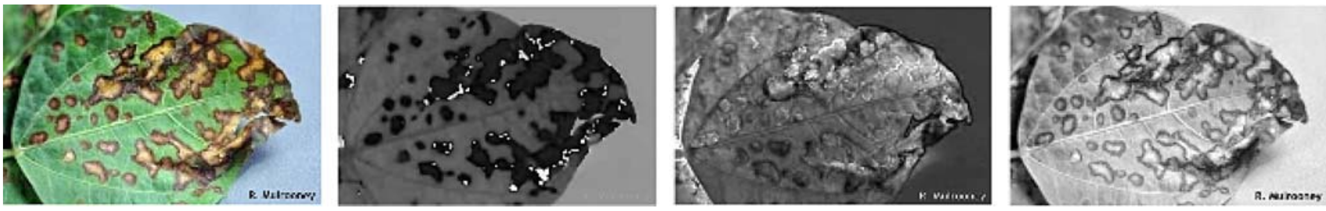
6) Obtain the useful segments;

7) Computing the texture features using Color-Co-Occurrence methodology;

8) Configuring the Neural Networks for Recognition.

2.1 Color transformation structure

First, the RGB images of leaves are converted into HSI color space representation. The purpose of the color space is to facilitate the specification of colors in some standard, generally accepted way. HSI (hue, saturation, intensity) color model is a popular color model because it is based on human perception (Gonzalez and Woods, 2008). Hue is a color attribute that refers to the dominant color as perceived by an observer. Saturation refers to the relative purity or the amount of white light added to hue and intensity refers to the amplitude of the light. Color spaces can be converted from one space to another easily. After the transformation process, the H component is taken into account for further analysis. S and I components are dropped since it does not give extra information. Figure 2 shows the H, S and I components.



a. Input image infected by bacterial brown spot b. Hue component c. Saturation component d. Intensity component

Figure 2 HSV components of a image infected by brown spots

2.1.1 Masking green pixels

In this step we identify the mostly green colored pixels. After that, based on specified threshold value that is computed for these pixels, the mostly green pixels are masked as follows: if the green component of the pixel intensity is less than the pre-computed threshold value, the red, green and blue components of the this pixel is assigned to a value of zero. This is done in sense that the green colored pixels mostly represent the healthy areas of the leaf and they do not add any valuable weight to disease identification. Furthermore this significantly reduces the processing time.

2.1.2 Removing the masked cells

In this step, the pixels with zeros red, green, blue values were completely removed. This is helpful as it gives more accurate disease classification and significantly reduces the processing time.

2.2 Segmentation:

From the above steps, the infected portion of the leaf is extracted. The infected region is then segmented into a number of patches of equal size. The size of the patch is chosen in such a way that the significant information is not lost. In this approach patch size of 32×32pixels is taken. The next step is to extract the useful segments. Not all segments contain significant amount of information. So the patches which are having more than fifty percent of the information are taken into account for the further analysis.

2.3 Color co-occurrence method

The color co-occurrence texture analysis method is developed through the SGDM. The gray level co-occurrence methodology is a statistical way to describe shape by statistically sampling the way certain gray-levels occur in relation to other gray levels (Argenti et al., 2008). These matrices measure the probability that a pixel at one particular gray level will occur at a distinct distance and orientation from any pixel given that pixel has a second particular gray level. The SGDM's are represented by the function $P(i, j, d, \theta)$ where i represent the gray level of the location (x, y), and j represents the gray level of the pixel at a distance d from location (x, y) at an orientation angle of θ. SGDM's are generated for H image.

2.4 Texture features

Texture features like Contrast, Energy, Local homogeneity, Cluster shade and cluster prominence are computed for the H image as given in Equations (1) to (5).

$$\text{Contrast} = \sum_{i,j=0}^{N-1} (i,j)^2 C(i,j) \tag{1}$$

$$\text{Energy} = \sum_{i,j=0}^{N-1} C(i,j)^2 \tag{2}$$

$$\text{Local Homogeneity} = \sum_{i,j=0}^{N-1} C(i,j) / (1+(i-j)^2) \tag{3}$$

$$\text{Cluster Shade} = \sum_{i,j=0}^{N-1} (i - M_x + j - M_y)^3 C(i,j) \tag{4}$$

$$\text{Cluster Prominence} = \sum_{i,j=0}^{N-1} (i - M_x + j - M_y)^4 C(i,j) \tag{5}$$

From the texture features, the plant diseases are classified into various types.

2.5 Classifier

2.5.1 Minimum distance criterion

In the classification phase, the co-occurrence features for the leaves are extracted and compared with the corresponding feature values stored in the feature library. The classification is first done using the Minimum Distance Criterion - (Arivazhagan et al., 2010). The success of classification is measured using the classification gain (G) and is calculated using Equation (6).

$$G(\%) = \frac{C_{corr}}{M} \times 100 \quad (6)$$

where, C_{corr} is the number of images correctly classified and M is the total number of images belonging to the particular texture group.

2.5.2 SVM classifier

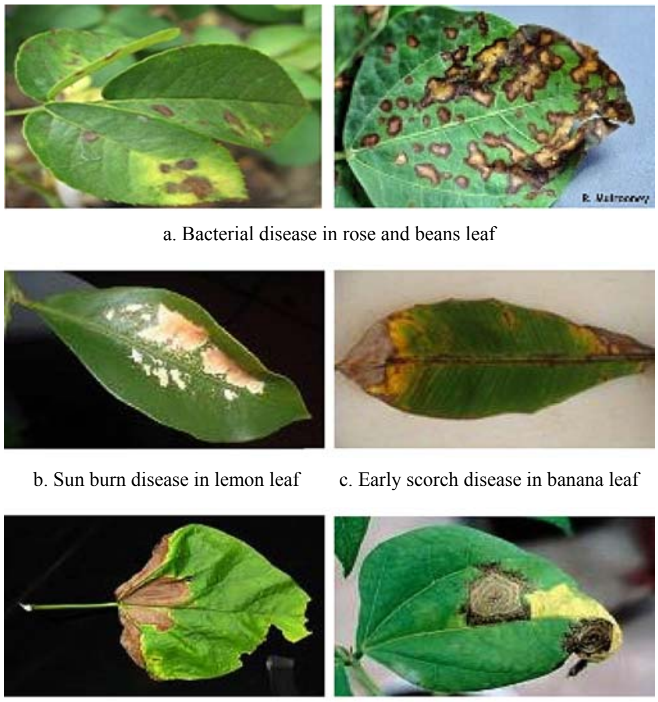
Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. Supervised learning involves analyzing a given set of labeled observations (the training set) so as to predict the labels of unlabelled future data (the test set). Specifically, the goal is to learn some function that describes the relationship between observations and their labels (Chi & Lin, 2002). More formally, a support vector machine constructs a hyper plane or set of hyper planes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), in general the larger the functional margin the lower the generalization error of the classifier.

In the case of support vector machines, a data point is viewed as a p-dimensional vector (a list of p numbers), and we want to know whether we can separate such points with a $(p-1)$-dimensional hyper plane. This is called a linear classifier. There are many hyper planes that might classify the data. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two classes. So we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized.

Multiclass SVM aims to assign labels to instances by using support vector machines, where the labels are drawn from a finite set of several elements. The dominant approach for doing so is to reduce the single multiclass problem into multiple binary classification problems. Common methods for such reduction include: building binary classifiers which distinguish between (i) one of the labels and the rest (*one-versus-all*) or (ii) between every pair of classes (*one-versus-one*). Classification of new instances for the one-versus-all case is done by a winner-takes-all strategy, in which the classifier with the highest output function assigns the class (Chamasemani and Singh, 2011).

3 Results and discussion

About 500 plant leaves of 30 different native plant species of Tamil Nadu have been collected for our approach. The acquired leaf images are converted into HSI format. The co-occurrence features like contrast, energy, local homogeneity, shade and prominence are derived from the co-occurrence matrix. With these set of co-occurrence features the plant diseases are detected. Samples of leaves with various diseases like early scorch, yellow spots, brown spots, late scorch, bacterial and fungal diseases are shown in Figure 3.



a. Bacterial disease in rose and beans leaf

b. Sun burn disease in lemon leaf

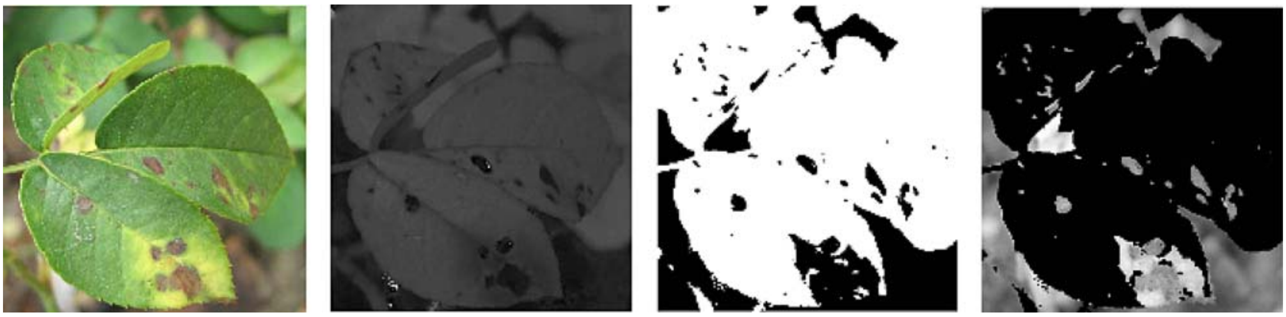
c. Early scorch disease in banana leaf

d. Late scorch disease in beans leaf

e. Fungal disease in beans leaf

Figure 3 Sample images of infected leaves

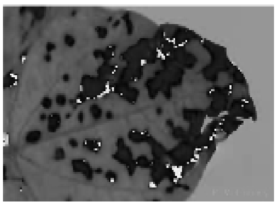
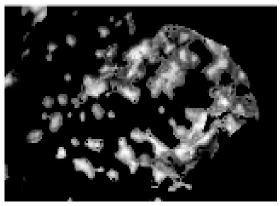
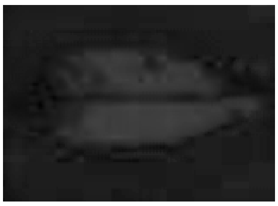
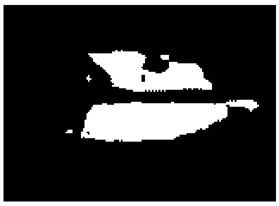
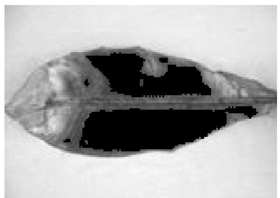
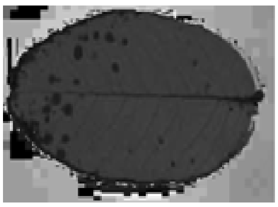
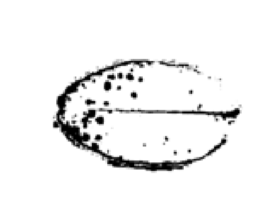
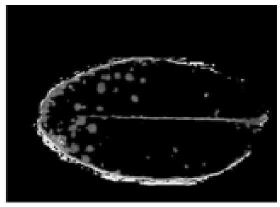
As a sample, a rose leaf that is infected by bacterial disease is given as input to the algorithm. Color transformation structure on the input image is performed. Then the green pixels are masked and removed using a specific threshold value. Then the R, G, B components are mapped to the thresholded image. These steps are shown in Figure 4. Table 1 lists the set of leaves that are affected by various diseases.



a. Input image b. Hue content c. Threshold image d. R component mapped output

Figure 4 Detection of infected region for a rose leaf

Table 1 Detected diseased region of various leaves

Plant species	Input image	Hue content	Thresholded image	R component mapped output
Beans				
Lemon				
Banana				
Guava				

After mapping the R, G, B components of the input image to the thresholded image, the co-occurrence features are calculated. The co-occurrence features for the leaves are extracted and compared with the corresponding feature values stored in the feature library. The classification is first done using the Minimum Distance Criterion. The leaf images are divided into training and testing set, where 5% of the leaf images from each group are used to train the system and the remaining images serves as the testing set. The number of images used for training, testing and classification gain for each type of leaves is shown in Table 2.

The classification gain obtained by Minimum Distance Criterion is 86.77%. The detection accuracy is improved to 94.74% by SVM classifier. The training and the testing sets for each type of leaf along with their detection accuracy is shown in Table 2.

From the results it can be seen that the detection accuracy is enhanced with SVM classifier.

The two class problem is then extended to multiclass problem where the detected leaf diseases are then classified into various categories. Training and the testing sets for each type of leaf disease along with their detection accuracy is shown in Table 3.

Table 2 Comparison of results by minimum distance classifier and support vector machine

Plant species	No. of images used for training	No. of images used for testing	Detection accuracy/%	
			MDC	SVM
Banana	10	10	82.25	90
Beans	10	12	96.43	91.66
Guava	10	14	78.95	92.86
Jackfruit	10	10	82.35	100
Lemon	10	30	94.4	96.66
Mango	10	17	80	94.12
Potato	10	10	78.57	100
Tomato	10	27	95.24	92.59
Overall accuracy			86.77	94.74

Table 3 Results of leaf disease recognition system

Plant Species	Category	No. of Images used for Training	No. of Images used for Testing	Detection Accuracy
Banana	Good	5	6	84.60%
	Late scorch	5	7	
Beans	Good	4	9	87.50%
	Bacterial spot	2	4	
	Fungal spot	4	11	
Guava	Good	5	7	92.86%
	Chocolate spot	5	7	
Jackfruit	Good	4	5	90%
	Bacterial disease	4	3	
	Fungal disease	2	2	
Lemon	Good	4	5	82.14%
	Bacterial disease	4	20	
	Sun burn	2	3	
Mango	Good	3	6	83.33%
	Bacterial disease	3	4	
	Sooty mold	4	8	
Potato	Good	4	7	96.43%
	Early blight	3	9	
	Late blight	3	12	
Sapota	Good	4	5	80%
	Scorch	3	3	
	Ashen mold	3	2	
Tomato	Good	4	5	82.15%
	Bacterial disease	4	19	
	Leaf lesion	2	4	
Overall Accuracy				87.66%

4 Conclusion

An application of texture analysis in detecting and classifying the plant leaf diseases has been explained in this paper. Thus the proposed algorithm was tested on ten species of plants namely banana, beans, jackfruit, lemon, mango, potato, tomato, and sapota. The diseases specific to those plants were taken for our approach. The experimental results indicate the proposed approach can recognize and classify the leaf diseases with a little computational effort. By this method, the plant diseases can be identified at the initial stage itself and the pest control tools can be used to solve pest problems while minimizing risks to people and the environment. The reasons for misclassification are as follows: the symptoms of the diseased plant leaves vary (at the beginning, tiny, dark brown to black spots, at later time, it has the phenomena of withered leaf, black or part leaf deletion), also the taken feature identification vectors need to further optimized. In order to improve disease identification rate at various stages, the training samples can be increased and shape feature and color feature along with the optimal features can be given as input condition of disease identification.

References

Al-Bashish, D., M. Braik, and S. Bani-Ahmad. 2011. Detection and classification of leaf diseases using K-means-based segmentation and neural networks based classification. *Information Technology Journal*, 10(2): 267-275.

Al-Hiary, H., S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. AlRahamneh. 2011. Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications*, 17(1): 31-38.

Argenti, F., L. Alparone, and G. Benelli. 1990. Fast algorithms for texture analysis using co-occurrence matrices. *IEEE proceedings*, 137, (6): 443-448.

Arivazhagan, S., R. N. Shebiah, S. S. Nidhyanandhan, and L. Ganesan. 2010. Classification of citrus and non-citrus fruits using texture features. *Computing Communication and Networking Technologies*, ICCCNT-2010.

Bauer, S. D., F. Korc, W. Forstner. 2011. The potential of automatic methods of classification to identify leaf diseases from multispectral images. *Precision Agriculture*, 12: 361-377.

Chamasemani, F. F., and Y. P. Singh. Malaysia, Bio-Inspired Computing: Theories and Applications (BIC-TA), 2011 Sixth International Conference. Faculty of Information Technology, MultiMedia University. Cyberjaya, Malaysia.

Chih-Wei, H., and C. Lin. 2002. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2): 415-425.

Helly, M. E., A. Rafea, and Salwa-El-Gammal. 2003. An integrated image processing system for leaf disease detection and diagnosis. in Proc. IICAI, pp.1182-1195.

Jean, W. 2009. Extension plant pathologist, georgia plant disease loss estimates, www.caes.uga.edu/publications.

Jun, W., and S. Wang. 2008. Image thresholding using weighted parzen window estimation. *Journal of Applied Sciences*, 8(5):772-779.

Kim, D. G., T. F. Burks, J. Qin, and D. M. Bulanon. 2009. Classification of grapefruit peel diseases using color texture feature analysis. *International Journal on Agriculture and Biological Engineering*, 2(3): 41-50.

Rastogi, R., and V. K. Chadda. Applications of image processing in biology and agriculture J. K. Sainis, Molecular Biology and Agriculture Division, BARC newsletter.

Siddiqil, M. H., S. Sulaiman, I. Faye, and I. Ahmad. 2009. A real time specific weed discrimination system using multi-level wavelet decomposition. *International Journal of Agriculture and Biology*, 11(5): 559-565.

Gonzalez, R., R. E. Woods. 2008 Digital image processing. Third edition, Pearson Education, Prentice-Hall, Inc.