Detection Of Plant Leaf Disease Using Image Processing Techniques

Bindushree H B, Dr. Sivasankari G G

AMCEC, Dept.CSE, Bangalore, India;
AMCEC, HOD Dept.CSE, Bangalore, India
bindushreebalasubramanya@gmail.com, mailtosankari@gmail.com

ABSTRACT: An estimated seventy percent of Indian economy depends on agriculture. Since there is growing Indian population, which is increasingly dependent on the agricultural yield, production of the crops must be enhanced. In order to grow more the diseases must be analysed in prior. Diseases are analysed using different image processing techniques. One such technique is proposed here. The proposed framework has been implemented in three phases. First, image segmentation is performed using K means clustering to identify the diseased regions area. In the next step features are extracted from segmented regions using feature extraction techniques such as GLCM. These feature are then used for classification into healthy or disease affected type. Experimental results of classification using Support Vector Machines show that our proposed technique is quite significantly better than any other existing techniques used for Plant Disease Detection and Identification and Support Vector Machines provides very accurate classification results.

1 INTRODUCTION

Plant diseases have turned into a major problem as it can cause significant reduction and losses in both quality and quantity of agricultural products [1]. A vast majority of the growing national population depends on agriculture yields. Farmers have wide range of diversity to select suitable fruit or vegetable crops to grow. But the cultivation of these crops [2] for optimum yield and quality produce is highly technical & challenging [3]. It can be improved by the aid of technological support and mechanized farming [4]. The management of perennial crops requires continuous and close monitoring [5] especially for the management of diseases that can affect production significantly [3]. Many authors have worked on the development methods for the automatic detection and classification of leaf diseases based on high resolution multispectral, hyperspectral and stereo images. The philosophy behind precision agriculture is not only including a direct economical optimization of agricultural production [6], it also stands for a reduction of harmful outputs into environment and non-target organisms [7]. In particular a contamination of water, soil, and food resources with pesticides has to be as minimal as possible in crop production. Automatic detection of plant diseases is a very important research topic as it may prove the benefits in monitoring large fields of crops, and thus automatically detect the symptoms of diseases [8] as soon as they appear on plant leaves. Therefore looking for fast, automatic, less expensive [4] and accurate method to detect plant disease cases is of great realistic significance [9]. Machine learning [10] based detection and recognition of plant diseases can provide extensive clues to identify [4] and treat the diseases in its very early stages. Comparatively, visually or naked eye identification of plant diseases is quite expensive, inefficient, inaccurate and difficult. Also, it requires the expertise of a well-trained botanist [10]. In [11] the authors have worked on the development of methods for the automatic classification of leaf diseases based on high resolution multispectral, hyperspectral and stereo images. Leaves of sugar beet are used for evaluating their approach. Sugar beet leaves might be infected by several diseases, such as rusts, powdery mildew. In [11], a fast and accurate new method is developed based on computer vision image processing for grading of plant diseases. For that, leaf region was segmented by using Otsu's [3] method. After that

the disease spot regions were segmented by using Sobel edge operator [12] to detect the disease spot edges. Finally, plant diseases are evaluated by calculating the quotient of disease spot and leaf areas. Previous works show that Machine learning methods can successfully be applied as an efficacious disease detection mechanism. Examples of machine learning methods that have been applied in agricultural researches are Artificial Neural Networks (ANNs), Decision Trees, K-means, k nearest neighbors, Support Vector Machines (SVMs) and BP Neural Networks. For example, Wang et al. in [1] predicted Phytophthora infestans [6] infection on tomatoes by using ANNs. Also, Camargo and Smith in [4] used SVMs to identify visual symptoms of cotton mould diseases using SVMs. There are two main important characteristics of plant disease detection machine-learning methods that must be investigated, they are: speed and accuracy. In this study an automatic detection and classification of leaf diseases is been proposed which is literally based on K-means as a clustering and SVM as classifier.

2 MACHINE LEARNING

A total of five different Machine Learning techniques for learning classifier have been investigated in this paper. These techniques are selected due to the reason that these classifiers have performed well in many real applications.

2.1 K- Nearest Neighbor (KNN)

The K Nearest Neighbor is a slow learner which means that this classifier can train and test at the same time. KNN classifier is an instance based classifier that performs classification of unknown instances by relating unknown to known by using distance or such similarity functions. It takes K nearest points and then assigns class of majority to the unknown instance [11].

2.2 Naïve Bayes Classifier

Naïve Bayesian Classification is commonly known as a statistical means [14] classifier. Its foundation is on Bayes' Theorem, and uses probabilistic analysis for efficient classification. Naïve Bayesian Classifier [14] give more accurate results in less computation time when applied to the large data sets consisting of hundreds of images.

2.3 Support Vector Machine (SVM)

Support Vector Machine is machine learning technique which is basically used for classification. It is a kernel based classifier; it was developed for linear separation which was able to classify data into two classes only. SVM has been used for different realistic problems such as face, gesture recognition [10], cancer diagnosis [8] voice identification and glaucoma diagnosis.

2.4 Decision Tree

An IJTEEE copyright form must accompany your final Decision Tree Classifiers (DTC's) are being successfully used in many areas including medical diagnosis, prognosis, speech recognition, character recognition etc. Decision tree classifiers have ability to convert the complex decision into easy and understandable decisions. [7]

2.5 Recurrent Neural Networks

Recurrent Neural Networks (RNN) includes feedback connections. In contrast to feed-forward and back propagation networks, the dynamical properties are more significant. Neural Network [6] has evolvement within a constant state and the activation values of any units do not change anymore. But in some cases, according to required scenario it is important change the activation value of the output neurons [6].
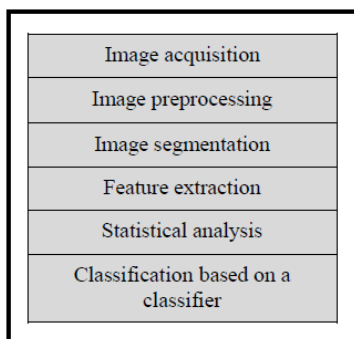
3 PROPOSED METHOD

Figure. 1 The proposed Method

3.1 Image Dataset

Data set is prepared and is used in this research. Leaf samples were acquired from indoor and natural scenes. This consists of two types of leaves. These leaves will be divided into normal and diseased. The Nikon camera D90, which is 15 megapixels camera, is used for image acquisition purpose. Program mode is used for more detail green and blues. Aperture is normal about 35 and the lens used is 18-135 mm. Distance of object from lens is about 9 to 12 inches. Images are acquired in indoor lighting.

3.2 Segmentation using K means clustering

K-means clustering is a partitioning/dividing method. The function kmeans partitions data into k mutually exclusive clusters [15], and returns the index value of the cluster to which it has assigned each observation. Unlike hierarchical clustering and semi hierarchical, k-means clustering operates on actual observations or rows (rather than the larger set of dissimilarity measures), and creates a single level of many clusters. The conclusions mean that k-means clustering is often more suitable than hierarchical clustering for large amounts of data. K

means treats each observation or row in the data as an object having a location in space. It finds a meaningful partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. You can choose from multiple distance measures, depending on the kind of data that we are clustering. Each cluster in the partition is defined by its member objects and by its centroid, or even the center. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized and evaluated. Kmeans computes cluster centroids differently [8] for each distance measure, to minimize the sum with respect to the measure that which has been specified. K means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters that have been created. This algorithm moves objects in between clusters until the sum cannot be decreased any further. The result is a set of clusters that are as compacted and are well-separated as far as possible. One can control the details of the minimization using several optional input parameters to kmeans clusters, including ones for the initial values of the cluster centroids, and for the maximum number of iterations. By default and by choice, kmeans uses the k-means algorithm for cluster center initialization and the squared Euclidean distance metric to determine distances. However, Kmeans clustering is used to partition the leaf image into three or more clusters in which one or more clusters contain the disease in case when the leaf is infected by more than one disease. In our implementation multiple values of number of clusters have been tested. Best results were obtained when the number of clusters was either 3 or 4.
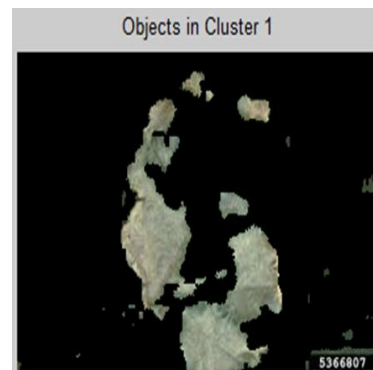
.

Figure 2 Disease affected leaf

Figure. 3 Disease affected area obtained by K means clustering

3.3 Feature Extraction

Statistical texture based features are extracted using Gray Level Cooccurance Matrix (GLCM). These are spatial features that indicate pixel relationship based on gray scale intensity and orientation. [19] A total of 11 Haralick features are used which are calculated using Gray Level Co-occurrence Matrix (GLCM). Table 1 shows the description of how each texture feature is calculated. In the equations, n represent the number of observed values. X is the sample space and P is the population.

Angular second moment(energy)	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}[p(i,j)]^2$	(1)
Corelation	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{ijp(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$	(2)
Variance	$\sigma^2 = \sum_{i=0}^{G-1}(i-\mu)^2 p(i)$	(3)
Inverse difference	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\frac{p(i,j)}{1+(i-j)^2}$	(4)
Sum Avg	$\sum_{n=0}^{G-1} n^2\, p_{x-y}(n)$	(5)
Sum variance	$\sum_{i=2}^{2N_g}(i-f6)^2\, p_{x+y}(i)$	(6)
Sum entropy	$\sum_{i=2}^{2N_g} p_{x+y}(i)log\big(p_{x+y}(i)\big)$	(7)
Entropy	$\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} p(i,j)\log_2[p(i)]$	(8)
Diff variance	$\sum_{i=0}^{N_{g-1}}(i-\mu_{x-y})^2\, p_{x-y}(i)$	(9)
Diff entropy	$\sum_{i=2}^{2N_g} p_{x-y}(i)log\big(p_{x-y}(i)\big)$	(10)
Information measure of correlation 1	$\left[\dfrac{f9-HXY1}{\max\{HX,HY\}}\right]$	(11)

Table. 1 Statistical Features

To create a GLCM, we make use of the graycomatrix function in Matlab. The graycomatrix function creates a new gray-level co-occurrence matrix (GLCM) [6] by calculating how often a pixel with the intensity (gray-scale -level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship [15] is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent and vertically seperated), but you can specify other spatial relationships [14] between the two pixels. Each element of pixels (i,j) in the resultant glcm is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the corresponding input image. The number of gray levels [15] in the digital image determines the size of the GLCM. By default, graycomatrix uses scaling of intensity to reduce the number of intensity values in an image to eight, but you can use the number of levels and the gray limits parameters to control this scaling of gray levels [15]. The gray-level co-occurrence matrix can reveal certain properties [14] about the spatial distribution of the gray levels in the texture image [17]. For example, if most of the entries in the GLCM are concentrated along the diagonal [17], the texture is coarse with respect to the specified offset

indicated. We have derived several statistical measures from the GLCM.

3.4 Classification

In supervised machine learning, support vector machines (SVMs, also support vector networks are supervised learning models with associated learning capable algorithms that analyze data and recognize patterns, used for binary classification and to plot regression analysis. Given a set of training samples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped such that the examples of the separate categories are divided into a clear gap that is as wide as possible. New samples are then mapped into that same space and are predicted to belong to a category based on which side of the gap which they fall on. In addition to performing linear binary classification, SVMs can efficiently perform a non-linear binary classification using what is called the kernel trick, which maps their inputs into high-dimensional feature spaces. The models are trained using svmtrain () and classified using svmclassify () commands in Matlab.The kernels used are:
* Linear
* Quadratic
* Polynomial
* MLP
* RBF

4 Results and Discussions

We have used MATLAB 2013 for the experimentation of the proposed system. Once the dataset is prepared, we segmented the images using K means clustering. Out of the three clusters created one of the clusters contain the disease affected area. Then we have extracted features from that particular cluster using Gray Level Cooccurance Matrix (GLCM). These features are later fed into Support Vector Machines (SVM). The final classification results from SVMs indicate whether the leaf in the image dataset is healthy or disease affected. The results using SVM are obtained from various kernels such as linear, polynomial, quadratic, RBF and polynomial. All the results from the kernels are able to predict the type of the image very accurately

REFERENCES

[1] Camargo A. and J. S. Smith. 2008. An image-processing based algorithm to automaticallyidentify plant disease Visual symptoms. Bio.Systematic. Engineering., 102: 9 – 21.

[2] Camargo, A. and J. S. Smith. 2009. Image processing for pattern classification for the identification of disease causing agents in plants. Com. Elect. Agr. 66: 121–125.

[3] Guru, D. S., P. B. Mallikarjuna and S. Manjunath. 2011. Segmentation and Classification of Tobacco Seedling Diseases. Proceedings of the Fourth Annual ACM Bangalore Conference.

[4] Zhao, Y. X., K. R. Wang, Z. Y. Bai, S. K. Li, R. Z. Xie and S. J. Gao. 2009. Research of Maize Leaf Disease

Identifying Models Based Image Recognition. Crop Modeling and Decision Support.Tsinghua uni.press. Beiging. pp. 317-324.

[5] Al-Hiaryy, H., S. Bani Yas Ahmad, M. Reyalat, M. Ahmed Braik and Z. AL Rahamnehiahh. 2011. Fast and Accurate Detection and Classification of Plant Diseases. Int. J. Com. App., 17(1): 31-38.

[6] PearlMutter, B. A. 1990. Dynamic Recurrent Neural Network Aly, M. 2005. Survey on Multiclass Classification Methods

[7] Fury, T. S., N. Cristianini and N. Duffy. 2000. Support vector machine (SVM) classification and validation of cancer tissue samples using microarray expression data. Proc. BioInfo., 16(10): 906-914.

[8] Scholkopf, B. and A. J. Smola. 2001. Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge.

[9] Huang, J., V. Blanz and B. Heisele. 2002. Face Recognition Using Component-Based SVM Classification and Morphable Models, pp. 334– 341.

[10] Mohammed J. Islam ., Q. M. Jonathan Wu, Maji-dAhmadi, A.Maher and Sid-Ahmed.2007. Investigating the Performance of Naive- Bayes Classifiers and K-Nearest Neighbor Classifiers.ICCI Proceedings of International Conference on Convergence Information Technology .IEEE Computer Society

[11] Bock, C. H., G. H. Poole, P. E. Parker and T. R. Gottwald. 2010. Plant Disease Severity Analysis Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral & Multispectral Imaging. Cri. Rev. Pla. Sci., 29: 59–107.

[12] http://users.rowan.edu/~polikar/WAVELETS/WTtuto rial.com (Accessed: 25th April 2013)

[13] Naveed N., T. S., Choi and A .Jaffa .Malignancy and Abnormality Detection of Mammograms using DWT features and ensembling of classifiers, International Journal of the Physical Sciences ,Vol.6(8)

[14] Duda, R. O., P. E. Hart and D. G. Stork. 2001. Pattern classification, 2nd edition, John Wiley and Sons, New York.

[15] Al Bashisha, D., M.Braika and S.BaniYas - Ahmad.2010. Frame work for detection and classification of plant leaf and stem diseases. Signal and Image Processing (ICSIP) international conference. pp.113 – 118

[16] Hongzhi, W., Ying, D, .2008. An Improved Image Segmentation Algorithm Based on Otsu Method. International Symposium on Photo electronic Detection and Imaging SPIE Vol. 6625

[17] Chung, K, L., Liu, Y, W, And Yan, W, M., 2006. A hybrid grey scale image representation using spatial – and DCT domain – based approach with application to moment computation. Journal of Visual Communication and Image Representation Vol. 17, Issue 6.

[18] Automated Plant Disease Analysis: Performance Comparison of Machine Learning Techniques Akhtar, Akram.; Khanum, Akram .; Khan, S.Akram.; Shaukat, Akram. Frontiers of Information Technology (FIT), 2013 11th International Conference on Frontiers of Information Technology.