Plant Disease Detection using New Geometric Features with Standard Data Mining Methods

Pathan Ayesha, Pokharkar Krutika, Hande Prarthana, Prof. C. S. Arya

***Abstract*— Plant identification belongs to domain of data mining. Images of leaves are usually used as the main element to distinguish a plant from another. Feature extraction is necessary for plant identification. In the literature, most plant recognition systems use classification method, which has been adapted or modified to face this type of application. For identifying plant leaf diseases Segmentation technique is used. Most of the available techniques for plant disease segmentation use gray scale value . In addition, the system deals with desert-specific challenges, such as, dust, infertile sandy soil, constant wind, very low humidity, and the extreme variations in diurnal and seasonal temperatures.**

***Index Terms*— Agriculture, Fuzzy logic, Control, Disease diagnosis, Critical view, Plant disease detection, Image segmentation**

I. INTRODUCTION

Plant identification is very important in areas such as medicine, botany and food sector. It is also important and interesting for ecologists, herbal doctors and general public. However, accurate plant recognition is a task that requires some botanical expertise. The development of plant recognition methods is an active research area in last years. Most of plant identification methods use leaf due to its attractive properties, for example, leaf shape vary between different species, and leaves can be collected throughout the year. Typically, some features are extracted from images of leaves, and then data mining methods are applied to recognize plants we introduce three new geometric features for plant identification. These capture the vertical and horizontal symmetry of leaf. In order to test the effectiveness of the proposal, we did the following experiment.

These features are the following: area convexity, perimeter convexity, circularity, aspect ratio and null area. Then, we added the three proposed features and measured the performance of the same classification methods. We found that the performance of classifiers is improved dramatically throughout the ages, earth exposes many pollution factors such as chemicals in water and soil, air pollution, sun exposure and plants. Human neglects the importance of preserving the environmental leading to the increase of the pollution over years. They are utilizing the resources without thinking how the earth will be like in the future. Moreover, the plants are very important and it's complete the life cycle of the earth. People are consuming the plants for foods, oxygen’s and other needs.

Early detection of plant diseases plays a very important role in the field of agriculture, especially for agricultural country like India. Detection of plant diseases till date is mainly by observation through naked eye, which leaves room for error. Detection of plant diseases are automated in certain parts of the world. For automatic detection of plant diseases, segmentation of the diseased spot is very essential. At present, there are several methods available for the segmentation purpose. Some common methods for segmentation are threshold based methods, boundary based methods, region based methods, morphology based methods etc.

The desert weather is hot, humid, and dusty in the summer and cold with occasional rain in winter. The desert climate has unique agricultural characteristics.

II. RELATED WORK

The extraction of features from leaf is one of the important steps in plant identification. Therefore, many researchers have proposed features that intent to capture the main characteristics of leaves. In authors propose three features which are invariant to scale, these are the following: aspect ratio, radius ratio and slice ratio. Authors tested the usefulness of their features with only one classification method, support vector machine. In the author designed and implemented their system to to control the performance using Short Message Service (SMS) of cell phones. Meaning that the motors performances depend on turning ON/Off remotely using mobile phone from any brand and also by sends message when it started or done its performance. This Project has been implemented in India and deal with it weather condition. The motors will turn off as soon as the farmers get an alarm about the single phasing. They used a GSM with a digital mobile telephone system and basically what it does is compresses data and then sends it down channels with two others stream user's data.

In most of the methods gray channel information of the image is used for segmentation purpose. Most of the techniques used for segmentation of gray level images can be extended for coloured images. It must be noted that the performance of coloured image segmentation largely depends on the colour space chosen. The selection of a particular colour space for the segmentation technique solely depends on the specific application. In the past few decades, agricultural applications using image processing and pattern recognition techniques have been attempted by various researchers. Different researchers have used different tools and techniques to identify plant diseases suitable for the application. The most common tools that are used for the detection, classification and quantification of plant diseases are thresholding, colour analysis [4], support vector machines [5], region growing [6], fuzzy logic [7] etc. In this paper, region growing algorithm based on automatic seed selection has been implemented for coloured images. The method was proposed by Frank Y. Shih and Shouxian Cheng [8] to segment colour images using seeded region growing technique.

- Availability: The system operates successfully at any point of time. It is requested in Kuwaits' gardens, or farms. Also, anyone can check the state of the farm remotely.
- Reliability: The user is able to access the measurement and see the values of the sensors. The system allows manual control.
- Maintainability: AgriSys can be easily upgrade by adding components with improved features
- Easy to use: The system is easy to use by people from all different age, system doesn't require previous knowledge to use.

III. SYSTEM ARCHITECTURE

We provide the guideline to the farmer for selecting the crop cultivation as per the weather present in that area. In that we guide to farmer for selecting crop cultivation as considering the weather also provide the list of the crops area wise requested from farmer. Once the farmer selects the crop for the cultivation then he does the booking for that particular seed and then when he purchase seed from the supplier then we provide guidelines for that particular crop by notification through sms.

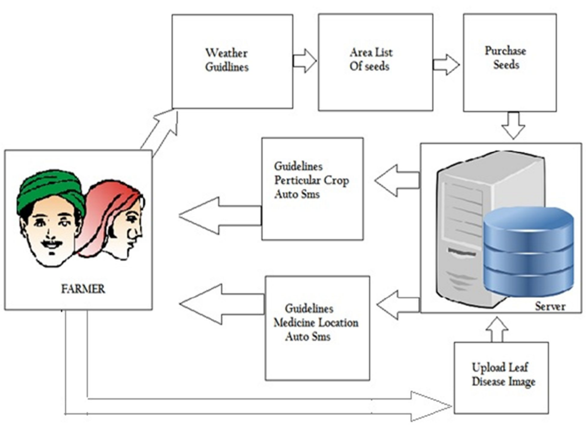


Fig.1. System Architecture

If some diseases occur on that crop then farmer simply capture the photo of the affected part and upload to our server then we provide the guidelines related to disease which has occurred on plant and also provide solution for that disease with the location of pestiside supplier.

We also provide area wise market price for that particular product.

IV. METHODOLOGY

The general process for plant identification consists of the following steps:

1) **Pre-processing:** Which includes Scaling, Normalization and Segmentation of images.
2) **Feature extraction:** Consists of obtaining geometric, contour or any other type of features from images of leave
3) **Classification:** Application of data mining methods sometimes adapted to this particular application to identify new leaves.

It is interesting to notice that the source code of the adaptations of classic classification methods for this application is not always publicly available. For this reason, it is important to develop new methods that allow applying directly data mining methods in this specific domain problem.

In the following subsections, we explain the general process to apply current classification methods for plant identification. A. Pre-processing and feature extraction The process for extracting the features from leaves requires that all images have the same size. However, most of leaf databases available on the Internet contain images with different dimensions. Therefore, we scale the images to a size of 1000_600 pixels. In order to extract useful features from a color image of a leaf, it is necessary to transform it into a high quality binary image. Normalization and segmentation are usually used to this purpose.

Image normalization: Used to have a certain independence of the properties of the image, such as brightness and contrast. This requires calculating the larger and smaller pixel values, max(I) and min(I), respectively. We use following equation to normalize the image:

N(x; y) = 255(I(x; y) ? min (I))

Leaf segmentation: We use a spatial threshold method to obtain a binary image of leaves. The threshold T(x; y) is calculated at each pixel, it depends on the probability distribution and/or correlation between pixels. The equation is used for this purpose.

Feature extraction : Feature extraction is one of the most important phases in plant recognition. There exist several well-known features for plant recognition.

V. ALGORITHM

1. SRG Algorithm For Colour Image Segmentation:

1. RGB colour image is transformed to YCbCr colour space model.

2. Initial seeds are selected automatically. Seed points should satisfy following conditions :

a) Seed pixels should have high similarity to its neighbours. Similarity of a pixel to its neighbourhood is calculated by following formula :

$$\sigma_x = \sqrt{\left(\frac{1}{9}\sum_{i=1}^{9}(x_i - \bar{x})^2\right)}$$

where x represents Y, Cb , or Cr , and x is the mean value of 3x3 neighbours.

Total standard deviation is calculated by :

$$\sigma = \sigma_Y + \sigma_{Cb} + \sigma_{Cr}$$

Similarity of a pixel of its 3x3 neighbourhood is calculated by:

$$H = 1 - \sigma_N$$

where σ_N is the normalized S.D. $\sigma_N = \sigma / \sigma_{max}$

To become eligible for seed pixel candidate, pixel should have H greater than some specific threshold value. Threshold is chosen by Otsu's Method

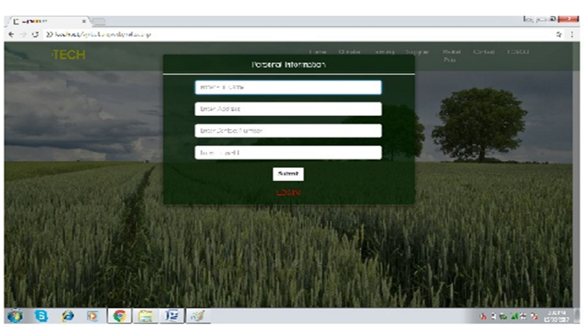
b) A seed pixel candidate must have the maximum relative Euclidean distance to its eight neighbors less than a threshold value which is chosen as 0.05 experimentally. Maximum relative Euclidian distance is calculated by following :

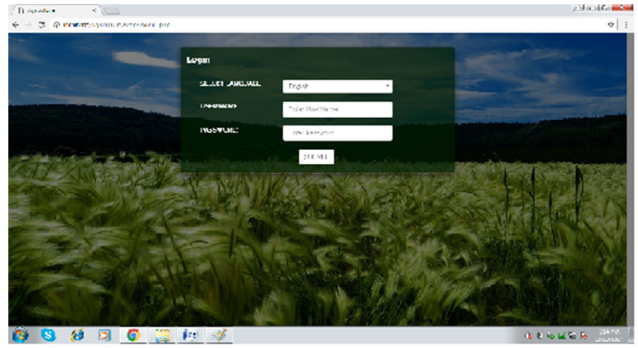
$$d_i = \frac{\sqrt{\left((Y-Y_i)^2 + (Cb-Cb_i)^2 + (Cr-Cr_i)^2\right)}}{\sqrt{\left(Y^2 + Cb^2 + Cr^2\right)}}$$

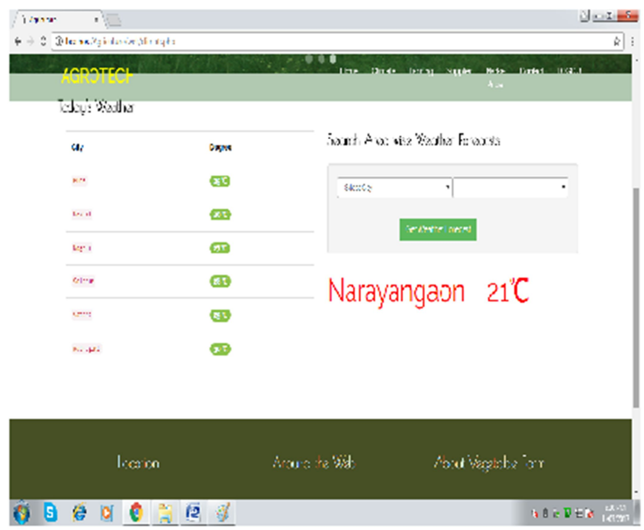
$$d_{max} = max(d_i) \text{ for i = 1 to 8}$$

3. Initial seeded regions are assigned different labels.
4. All four connected neighbours of initial seeds are stored in a sequentially sorted list (SSL) in increasing order of distance. Distance is measured by Euclidian distance metric.
5. If SSL is not empty following steps are done :
a) First point (P) of SSL is removed.
b) Labels of all four connected neighbours are checked.
c) If all neighbours of four connected regions have same label then set point P to this label.
d) Otherwise, Classify P to neighbouring region from which it has minimum distance.
e) Mean of the last labeled region is updated.
f) Add four connected neighbours of P which are neither labeled nor in the SSL to SSL.
6. Region merging is performed by following steps :
a) If mean colour difference between two regions d(Ri,Rj) < 0.1, then those two regions are merged. Those two regions for which mean colour difference is minimum are merged first.
b) If no. of pixels in a region < (1/150) * Total no. of pixels.
c) Then the region is merged into its nearing region with smallest colour difference.

VI. RESULTS







VII. CONCLUSION

A The development of methods for plant recognition is an active research area in last years. Recently, many researchers have proposed features to characterize leaf of plants, and use these features with classification methods to design plant recognition systems. However, most of such systems need to adapt such methods to make them suitable for this specific type of application. A drawback of this approach is that the source codes are not always available.

ACKNOWLEDGMENT

Authors thanks project guide Prof. C. S. Arya who always being with presence & constant, constructive criticism to made this paper. We would also like to thank all the staff of computer department for their valuable guidance, suggestion and support through the project work, who has given co-operation for the paper with personal attention. Above all we express our deepest gratitude to all of them for their kind-hearted support which helped us a lot during Paper work. At the last we thankful to our friends, colleagues for the inspirational help provided to us through a paper work.

REFERENCES

[1] G. Cerutti, L. Tougne, D. Coquin, and A. Vacavant, “Leaf margins as sequences: A structural approach to leaf identification,” Pattern Recognition Letters, vol. 49, pp. 177 – 184, 2014.

[2] J. S. Cope, D. Corney, J. Y. Clark, P. Remagnino, and P. Wilkin, “Plant species identification using digital morphometrics: A review,” Expert Systems with Applications, vol. 39, no. 8, pp. 7562 – 7573, 2012.

[3] C.-Y. Gwo, C.-H. Wei, and Y. Li, “Rotary matching of edge features for leaf recognition,” Computers and Electronics in Agriculture, vol. 91, pp. 124 – 134, 2013.

[4] C. Zhao, S. S. Chan, W.-K. Cham, and L. Chu, “Plant identification using leaf shapes-a pattern counting approach,” Pattern Recognition, vol. 48, no. 10, pp. 3203 – 3215, 2015, discriminative Feature Learning from Big Data for Visual Recognition.

[5] S. Watcharabutsarakham, W. Sinthupinyo, and K. Kiratiratanapruk, “Leaf classification using structure features and support vector machines,” in Information Science and Service Science and Data Mining (ISSDM), 2012 6th International Conference on New Trends in. IEEE, 2012.

[6] J. I. Arribas, G. V. S´anchez-Ferrero, G. Ruiz-Ruiz, and J. G´omez- Gil, “Leaf classification in sunflower crops by computer vision and neural networks,” Computers and Electronics in Agriculture, vol. 78, no. 1, pp. 9 – 18, 2011

[7] T. Munisami, M. Ramsurn, S. Kishnah, and S. Pudaruth, “Plant leaf recognition using shape features and colour histogram with k-nearest neighbour classifiers,” Procedia Computer Science, vol. 58, pp. 740 – 747, 2015, second International Symposium on Computer Vision and the Internet (VisionNet15).

[8] C. Uluturk and A. Ugur, “Recognition of leaves based on morphological features derived from two half-regions,” in Innovations in Intelligent Systems and Applications (INISTA), 2012 International Symposium on, July 2012, pp. 1–4.

[9] Wenjiang Huang, Qingsong Guan, Juhua Luo, Jingcheng Zhang, Jinling Zhao, Dong Liang, Linsheng Huang, and Dongyan Zhang, “New Optimized Spectral Indices for Identifying and Monitoring Winter Wheat Diseases”, IEEE journal of selected topics in applied earth observation and remote sensing,Vol. 7, No. 6, June 2014.

[10] Mrunalini R. Badnakhe, Prashant R. Deshmukh, “Infected Leaf Analysis and Comparison by Otsu Threshold and k-Means Clustering”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 3, March 2012.

[11] R.M. Haralick and L.G. Shapiro, "Image segmentation techniques,"Comput. Vis. Graph Image Process., vol. 29, pp. 100-132, 1985.

[12] Jayme Garcia Arnal Barbedo, "Digital image processing techniques for detecting, quantifying and classifying plant diseases," Journal on Springerplus, vol.2:660, 2013.

[13] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik and Z. ALRahamneh, “Fast and Accurate Detection and Classification of Plant Diseases”, International Journal of Computer Applications (0975 – 8887)Volume 17– No.1, March 2011