ISSN:0975-9646



# Uncovering of Plant Disease Based on Content Recovery from Images

Tejaswini C. Karpe[1], Deipali V. Gore[2]

[1]*Computer Department, P.E.S. Modern College of Engineering, Pune, India*

[2]*P.E.S. Modern College of Engineering, Pune, India*

***Abstract*—Web content mining is the process of pulling knowledge from the content of documents or their descriptions. This is one of the types of Web mining. Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence (AI), machine learning, statistics, and database systems. The term used earlier – 'Web mining' is one of the methods of data mining. It is the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the World Wide Web (WWW). This theory can be applied at various places like agriculture sector, etc.**

***Keywords*—Web mining, Support Vector Machine, AI, SGDM, Hue, SIFT, neural network classifier, DWT.**

## I. INTRODUCTION

Web Content Mining is the process of extracting useful as well as important information from the contents of Web documents. Web Content Mining is related to Data Mining because many Data Mining techniques and methods can be applied in Web Content Mining. It is also concerned with text mining since much of the web contents are text, but is also quite dissimilar from these because web data is mainly semi structured in nature and text mining focuses on formless text [1].

Web content mining is also different from text mining due to the semi-structure quality of the web, while text mining concentrates on amorphous texts. The methods that are generally used in web content mining are NLP (Natural language processing) and IR (Information retrieval)[2]. Web content mining is the scanning and mining of text, images and graphs of web page to determine relevance of content to the search query.

In this paper we are going to discuss some well known methods of Content recovery from images. In addition to this we are going compare and analyse these methods by results and outcome.

## II. LITERATURE SURVEY

Basavaraj S. Anami, Suvarna S. Nandyal, A. Govardhan explains a method for identification and classification of images of medicinal plants based on color canon and texture feature using SVM [3] and neural network classifier [3].

N.Valliammal Dr.S.N.Geethalakshmi [4] focuses on the pre-processing technique for Computer aided plant classification through leaf recognition.

Arunkumar Beyyala Sai Priya Beyyala [5] stressed on knowing the techniques for detection of plant traits or diseases using Image Processing. The main aim is to develop an effective image processing module for early diagnosis of disease, even before symptoms expression, for deadly diseases.

B.Sathya, Bamas.Mohana Valli, S.Raju, V. Abhai Kumar proposes an efficient computer-aided Plant Image Retrieval method based on plant leaf images using Shape, Color and Texture features intended mainly for medical industry, botanical gardening and cosmetic industry[6]. Bama and team used HSV color space to extract the various features of leaves. In addition to this log-Gabor wavelet is applied to the input image for texture feature extraction.

S. Ananthi,S. Vishnu Varthini proposed system is a software solution for automatic detection and computation of texture statistics for plant leaf diseases. They developed one 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, then the graphic is segmented & the utile segments are extracted, finally the texture statistics is computed. Main goal is to evaluate the presence of diseases on the plant leaf from the texture statistics[7].

H. Al-Hiary, S. Bani-Ahmad,M. Reyalat, M. Braik ,Z. ALRahamneh implemented software solution for automatic detection and classification of plant leaf diseases[8].

Ravikant Sinha, Pragya Pandey used Color texture feature analysis is used for detection of surface defects on pomegranates (Name of the tree). Acquired image is initially cropped and then transformed into HSI color space, which is further practiced for generating SGDM matrix [9]. Total 18 texture features were computed for hue (H), saturation (S) and intensity (I) images from each cropped samples. Best features were utilised as an input to Support Vector Machine (SVM) classifier and tests were performed to identify best classification model. An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

## III. CONTENT RECOVERY TECHNIQUES

Following methods are used by different researchers for content recovery from images.

- Support Vector Machine (SVM)
- Color Histogram
- Edge Histogram (EH)
- Edge Direction Histogram Texture

- Contrast Stretching Technique
- Adaptive Thresholding Method
- HSV Color Space
- The Scale Invariant Feature Transform (SIFT)
- K-Means Clustering
- Color Co-Occurrence Matrix
- Discrete Wavelet Transform (DWT)

In this section we are going to discuss below three methods.

A. Approach based on Color, texture and edge features
B. Pre-processing technique
C. Automatic detection and computation of texture statistics.

*A. Approach based on Color, texture and edge feature*

This method explains edge and color descriptors that have low-dimension, effective and simple. Furthermore, the revolution invariant texture descriptors namely, directional difference and the gradient histogram are used. Author performed this idea on Images taken from plants like Tulsi, Papaya, Aloe Vera etc.

Author used a methodology which gives the classification of images of medicinal plants by Support Vector Machine (SVM) [3] and neural network classifier [3] using color and texture. Proposed method has been depicted in below figure.

To find the features, namely, color histogram, edge histogram, edge direction histograms are extracted using Sobel operator [3].

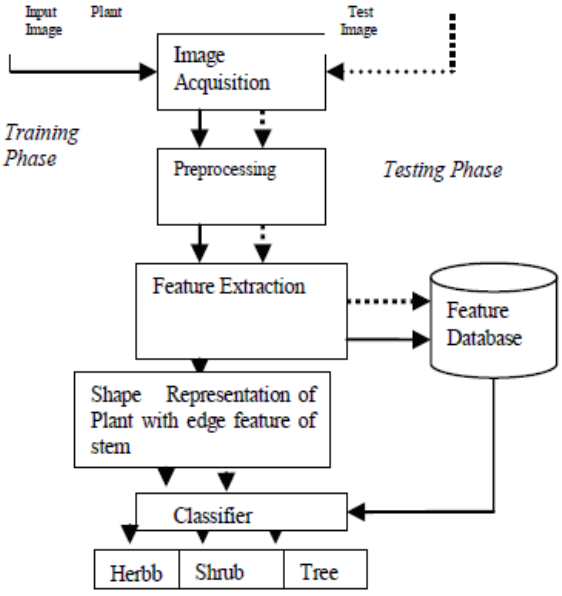


**Figure 1 Proposed Methodology**

- Color Histogram - Color histogram helps for finding color distribution of pixels in a plant image.
- Edge Histogram (EH) [3] – It is used for extracting textural features of stem and leaf portion of a medicinal plant. It is also useful in image matching in the absence of any homogeneous texture.
- Edge Direction Histogram Texture [3] - The Sobel operator helps in calculating the edges and direction of edges

*B. Pre-processing Techniques*

Pre-processing is the basic step to reconstruct the image with some useful feature. This method is crucial for the enhancement of leaf images which increases the efficiency of the subsequent tasks of the leaf recognition system.

A hybrid approach is proposed which is a combination of contrast stretching and adaptive thresholding that simultaneously adjusts the intensity level of leaf images using boundaries is developed.

Leaf images normally changes to blurred images by the existence of noise, low or high contrast both in the edge area and image area. Pre-processing an image include, removal of noise, edge or boundary enhancement, automatic edge detection, automatic contrast adjustment and segmentation. As multiple noises reduce the tone of nature images, improved enhancement technique is required for improving the contrast stretch in leaf images.

Image enhancement processes comprise of an accumulation of techniques that seek to improve the visual appearance of an image or to convert the image to a form better suited for analysis by a human or machine.

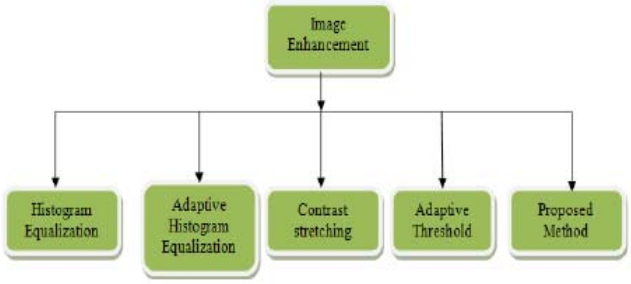


**Figure 2 Enhancement Techniques**

Out of above specified methods, used below techniques only.

1. **Contrast Stretching Technique** - The basic idea behind contrast stretching [4] is to linearly increase or decrease the contrast of the given image. This can be achieved by specifying the input/output relationship.
2. **Adaptive Thresholding Method** [4] - Thresholding is called adaptive thresholding when different thresholds are used for different regions in the image. This is also called as local or dynamic thresholding.

*C. Automatic detection and computation of texture statistics*

They developed one processing scheme consists of four important steps, first a colour transformation structure for the input RGB image is created, later the green pixels are masked and took off using specific threshold value, later the image is segmented and the useful segments are pulled, finally the texture statistics is computed.

Main goal is to evaluate the presence of diseases on the plant leaf from the texture statistics.

From the figure the factors are:-

**Color Transformation Structure** [7]: First, the RGB images of leaves are converted into Hue Saturation Intensity (HSI) color space representation. Color spaces can be converted from one space to another easily. Later on the

transformation process, the H component is taken into consideration for further analysis. S and I are dropped since it

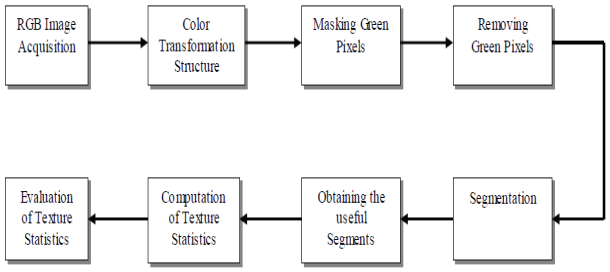


**Figure 3 Proposed approach**

Consideration for further analysis. S and I are dropped since it does not provide extra information.

**Masking green pixels**: In this step, we distinguish the mostly green colored pixels. if the green component of the pixel intensity is less than the pre-computed threshold value, the red, green and blue portions of the this pixel is assigned to a value of zero.

**Removing the masked cells**: The pixels with zeros red, green, blue components were completely moved out. This is helpful as it gives more accurate disease classification and significantly reduces the processing time.

**Segmentation** [7]: From the above steps, the septic portion of the leaf is extracted. The tainted region is then segmented into a number of patches of equal size.

**Color co-occurrence**: The color co-occurrence texture analysis method is developed through the Spatial Gray-level Dependence Matrices (SGDM) [7].

**Texture Features**: like Contrast, Energy, Local homogeneity, Cluster shade and Cluster prominence are computed for the Hue content of the image

## IV COMPARATIVE STUDY OF ABOVE METHODS

| Techniques | | | Result/Conclusion |
|---|---|---|---|
| 1 | 2 | 3 | |
| Color histogram | Edge Histogram | Histogram Texture | Good results are obtained by the combination of color and texture (edge) features. It is found that classification accuracy is better with SVM classifier than neural network classifier |
| Contrast Stretching Technique | Adaptive thresholding method | | As multiple noise damages the quality of nature images, improved enhancement technique is required for improving the contrast stretch in leaf images. |
| SGDM matrix | Support Vector Machine (SVM) | Color Co-Occurrence Matrix | This algorithm can be implemented for automatic grading and sorting system for quality control. |

## V. CONCLUSION

From above, it is clear that, fetching the content of the images including its color, sharpness, contrast, sheds, edges are very important in several areas like detecting plant diseases, medicals etc

In addition to this, it is very important to carry out the experiments on clear and exact images to attain the better result.

Most of time we faced the issue of long computational time along with difficulty working with blurred images.

## ACKNOWLEDGMENT

I am thankful to Prof. Ms. D. V. Gore for being my mentor in this journey. She has been the corner stone of my preparation and has guided me during my troubles.

I am also thankful to all the professors of our department for being a constant source of inspiration and motivation during the course of the Seminar.

We offer our heartiest thanks to our friends for their constant inspirations. Last but not the least, we want to acknowledge the contributions of our parents and family members, for their constant motivation, inspirations and their belief in us that we could make it happen.

## REFERENCES

[1] "Web Usage Mining: A Survey On Pattern Extraction From Web Logs" By S. K. Pani, Panigrahy, V.H.Sankar, Bikram Keshari Ratha, A.K.Mandal, S.K.Padhi P.G. Department Of Computer Science, RCMA; Bhubaneswar, Orissa, India.

[2] "WEB CONTENT MINING TECHNIQUES – A COMPREHENSIVE SURVEY" By Niki R. Kapadia And Kinjal Patel (ISSN: 2249-3905) International Journal Of Research In Engineering & Applied Sciences

[3] "A Combined Color, Texture And Edge Features Based Approach For Identification And Classification Of Indian Medicinal Plants" By Basavaraj S. Anami, Suvarna S. Nandyal, And A. Govardhan.

[4] "A Hybrid Method For Enhancement Of Plant Leaf Recognition" By N.Valliammal And Dr.S.N.Geethalakshmi.

[5] "APPLICATION FOR DIAGNOSIS OF DISEASES IN CROPS USING IMAGE PROCESSING" By Arunkumar Beyyala And Sai Priya Beyyala.

[6] "Content Based Leaf Image Retrieval (Cblir) Using Shape, Color And Texture Features" By B.Sathya Bama, S.Mohana Valli, S.Raju And V.Abhai Kumar

[7] "DETECTION AND CLASSIFICATION OF PLANT LEAF DISEASES" By S. Ananthi And S. Vishnu Varthini

[8] "Fast And Accurate Detection And Classification Of Plant Diseases" By H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik And Z. Alrahamneh

[9] "Infection Analysis Using Colour Feature Texture Using Image Processing" By Ravikant Sinha And Pragya Pandey

[10] "Leaf Image Segmentation Based On The Combination Of Wavelet Transform And K Means Clustering" By N.Valliammal And Dr.S.N.Geethalakshmi