IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

Detection of diseases on Cotton Leaves in Rural India Using CBIR

Bhagya Shri Mhetre[1], Reshma Bhosale[2,] Priyanka Kalme[3], Komal Lodhe[4], R.M. Samant[5]

Department of Information Technology, NBN Sinhgad School of Engineering, Pune, Maharashtra, India[1,2,3,4]

Professor, Department of Information Technology, NBN Sinhgad School of Engineering, Pune, Maharashtra, India[5]

Abstract: The main source for the disease is the leaf of the cotton plant. About 80 to 90 % of disease on the cotton plant is on its leaves. So for that our study of interest is the leaf of the cotton tree rather than whole cotton plant the cotton leaf is mainly suffered from diseases like fungus, foliar leaf spot of cotton, Alternaria leaf spot of cotton. The machine vision system now a day is normally consists of computer, digital camera and application software. We wanted to propose this project is to produce a working system that is capable of image retrieval through the use of content analysis and associated metadata in the form of supplied cotton leaf images. Existing system based on fuzzy feature selection approach using fuzzy curves (FC) and surface (FS). The reported accuracy in the literature is around 60%. We are using two approaches to image retrieval: Text-Based approach and Content-Based approach. Our approach is more effective and raises the accuracy rate to 80%.CBIR systems search collection of images based on features that can be extracted from the image Files themselves without manual descriptive. Our propose system will have android application for farmer which will allow farmer to take snapshot of cotton leaf and show the diagnosis.

Keywords: Fuzzy curves(FC),Fuzzy Surface(FS),Contain Based Image Retrieval(CBIR),Colour Co-occurrence Matrix(CCM),Gray Level Co-occurrence Matrix(GLCM),Histogram(HSV).

I. INTRODUCTION

There are two approaches to image retrieval: Text-Based approach and Content- Based approach. Today, the most common way of doing this is by textual descriptions and categorizing of images. This approach has some obvious shortcomings. Different people might categorize or describe the same image differently, leading to problems retrieving it again. It is also time consuming when dealing with very large databases. Content based image retrieval (CBIR) is a way to get around these problems.

The extraction of the features and image disease classification during these steps is as shown in the following figure.

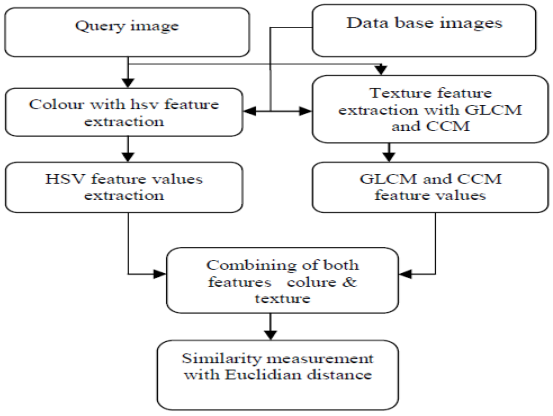

Fig 1. Process of CBIR system

1.1 Color feature extraction

HSV color model forms a uniform color space, which uses a linear gauge. The perceived distance between colors is in proportion to Euclidean distance between corresponding pixels in HSV color model, and conforms to eye's feeling about color. So it is very suitable for color based image similarity comparison.

1.2 Texture feature extraction based on GLCM

GLCM creates a matrix with the directions and distances between pixels, and then extracts meaningful statistics from the matrix as texture features. GLCM expresses the texture feature according the correlation of the couple pixels gray-level at different positions. It quantification ally describes the texture feature.
In this paper, four features is selected, include energy, contrast, entropy, inverse difference

Algo: GLCM
Input: Image File
Output: Energy, Entropy etc.
Steps:

1. resize the image as per database images size
2. Converts RGB values to grayscale values by forming a weighted sum of the R, G, and B components: 0.2989 * R + 0.5870 * G + 0.1140 * B
3. Create an empty GLCM matrix

4. For every pixel:
4.1. Calculate how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j.
4.2. Sum the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

4.3. Update GLCM matrix [i,j] with above sum value
4.4. any more pixels, yes continue, no break

5. After GLCM created, calculate:
5.1: Contrast - Measures the local variations in the gray-level co-occurrence matrix.
5.2: Correlation - Measures the joint probability occurrence of the specified pixel pairs.
5.3: Energy - Provides the sum of squared elements in the GLCM. Also known as uniformity or the angular second moment.
5.4: Homogeneity - Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

II. LITERATURE SURVEY

Various papers are suggesting to diagnosis the cotton leaves using various approach suggesting the various implementation ways as illustrated and discussed below. In the research of identifying and diagnosing cotton disease using computer vision intellectively in the agriculture, feature selection is a key question in pattern recognition and affects the design and performance of the classifier. In some paper, the fuzzy feature selection approach fuzzy curves (FC) and surfaces (FS) - is proposed to select features of cotton disease leaves image.

In order to get best information for diagnosing and identifying, a subset of independent significant features is identified exploiting the fuzzy feature selection approach. Firstly, utilize FC to automatically and quickly isolate a small set of significant features from the set of original features according to their significance and eliminate spurious features; then, use FS to get rid of the features dependent on the significant features. This approach reduces the dimensionality of the feature space so that lead to a simplified classification scheme appropriate for practical classification applications. The results show that the effectiveness of features selected by the FC and FS method is much better than that selected by human randomly or other methods. Also another approach is used to diagnosis the grape leaf disease identification or diagnosis, i.e. paper explaining the grape leaf disease detection from color imaginary using hybrid intelligent system, in that automatic plant disease diagnosis using multiple artificial intelligent techniques. The system can diagnose plant leaf disease without maintaining any expertise once the system is trained. Mainly, the cotton leaves disease is focused in this work. The proposed system consists of three main parts (a) cotton leaf color segmentation (b) cotton leaves disease segmentation and (c) analysis and classification of diseases.

Segmentation algorithms fall into two general classes, based on whether they searching for discontinuities or similarities. Algorithms focusing on locating discontinuities in the data are primarily edge-based, while algorithms concerned with locating adjacent pixels based on similarities are primarily region-based. Threshold techniques, a major category of algorithms, can fall into either class. In addition to these two major classes, there are also a number of general subcategories. For instance, algorithms either process color or gray-scale data, operate on either an individual pixel basis (global) or a neighbourhood of pixels (local), and may use different window sizes or different color representations. For example survey of segmentation algorithms. Cheng discussed the major segmentation approaches for segmenting monochrome images: histogram threshold, characteristic feature clustering, edge detection, region-based methods, fuzzy techniques, neural networks training method. The cotton leaf disease segmentation is performed using modified self-organizing feature map with genetic algorithms for optimization and support vector machines for classification.

Finally, the resulting segmented image is filtered by Gabor wavelet which allows the system to analyse leaf disease color features more efficient. The support vector machines are then again applied to classify types of grape leaf diseases. Similar idea can be extracted from to grape leaf disease diagnosis system and applicable to cotton leaves diagnosis system.

III. PROPOSED METHODOLOGY

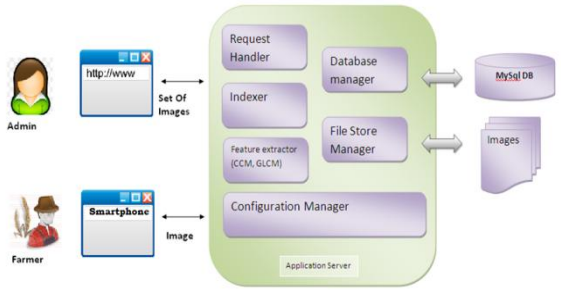

Fig 2: Architectural Design of CBIR System

3.1 Web based GUI

Server will be web based application and this module will be responsible to take inputs from admin. The GUI will be developed in HTML and Java-script

3.2 Android Client

An android application will be needed for farmer to provide input and to see the output on his device. Farmer will send image of the plant leaf to server. This module will take care of camera capturing activity that needs to be perform to take image.

3.3 Colour Feature Extraction

The most commonly used method to represent color feature of an image is the color histogram. A color histogram is a type of bar graph, where the height of each bar represents an amount of particular color of the color space being used in the image. The bars in a color histogram are named as bins and they represent the x-axis. The number of bins depends on the number of colors there are in an image. The number of pixels in each bin denotes

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 4, April 2016

y-axis, which shows how many pixels in an image are of a particular color. A three dimensional RGB (8*8*8) histogram contains total 512 bins. At the time of image indexing, the color of each pixel is find out, and its corresponding bin's count incremented by one.

3.4 Texture Feature Extraction

Texture feature extraction based on GLCM. A GLCM is a histogram of co-occurring grayscale values at a given offset over an image. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix.

3.5 Matching Logic

Euclidean Distance method is used to find out distance measures. These distance measurements indicated the similarities. The low value of distance measurement represent the close (good) similarity relation otherwise the high value of distance measurement represent the open (bad) similarity relation between two images. The metric used in Euclidean distance is called Euclidean metric to find out the distance measurements.

3.6 Database Manager

This module will help to handle all database related activity. All the SQL queries will be taken care in this module. A database connection polling system will be present to avoid repeatedly opening and closing database connection.

IV. RESULT AND ANALYSIS

4.1 Disease Analysis

We use following Cotton leaf Diseases in database

Disease Name	Images	Accuracy in %
Bacterial Blight	10	85
Leaf Spot	12	80
Leaf curl	15	84
White Files	10	87
Leaf Insect	8	85
Potassium Deficiency	3	78
Nitrogen Deficiency	2	75
Phosphorus deficiency	2	70
Magnesium deficiency	4	86
Zinc deficiency	5	80
Boron deficiency	6	82

Table No.1: Diseases name and Accuracy

4.2 Database image

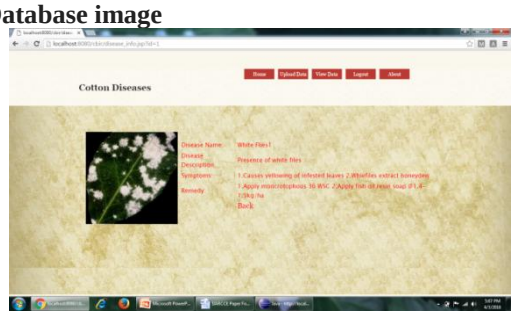

Fig.3: Database Image

4.3 Test image

Fig.4: user uploaded image

We use following concept for finding the result

Color

Computing distance measures based on color similarity is achieved by computing a color histogram for each image that identifies the proportion of pixels within an image holding specific values..

Texture

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image.

4.4 Derived value of database image and test image after implementation of Algorithm.

Database Image value:
Colors Expectancy: RED#62.638336181640625, GREEN#68.85832214355469, BLUE#51.25535583496094
Colors Variance: RED#77.24395272441078, GREEN#71.86810338869529, BLUE#72.494973224944
Skewness: RED#81.28335415388985, GREEN#71.62428373460111, BLUE#78.3886619126583
Test Image value:
Colors Expectancy: RED#64.48785400390625, GREEN#74.57963562011719, BLUE#56.20916748046875
Colors Variance: RED#75.12511621364852, GREEN#79.86765274679901, BLUE#73.97184060233283
Skewness: RED#71.15816832855621, GREEN#64.3114829473727, BLUE#78.55639791020666

RESULT

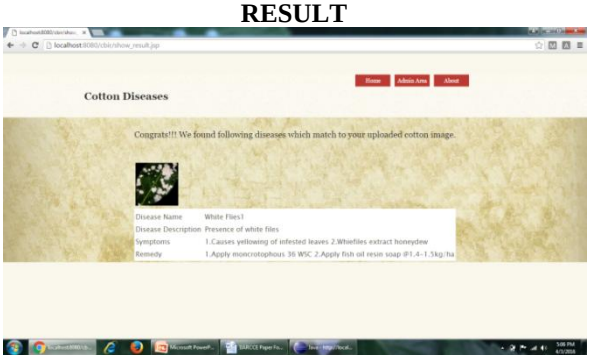

Fig 5: Result

V. CONCLUSION

In this paper, we have proposed an efficient CBIR (Content Based Image Retrieval) technique to retrieval cotton disease. Content based image retrieval system play an important role in matching the image having similar features. The main objective of this system was to match the query image with exiting images to see if they match in any features. Images are stored in database and the database schema is also designed for the system. The query sample image is also stored in the database now similar features like skewness, variance, HSV, color-expectancy are matched respectively.

REFERENCES

[1]. RitendraDatta, Dhiraj Joshi, Jia Li and James Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November 10-11, 2005, Hilton, Singapore.

[2]. C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," in IEEE Trans. On PAMI, vol. 24, No.8, pp. 1026-1038, 2002.

[3]. Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," in IEEE Trans. on PAMI, vol. 24, No.9, pp. 1252-1267, 2002.

[4]. A. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," in Proc. ACM SIGMOD Int. Conf. Management of Data, pp. 395–406, 1999.

[5]. J. Li, J.Z. Wang, and G. Wiederhold, "IRM: Integrated Region Matching for Image Retrieval," in Proc. of the 8th ACM Int. Conf. on Multimedia, pp. 147-156, Oct. 2000.

[6]. V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Region-based Image Retrieval Using an.