Tulsi Leaves Classification System

Kavya Bharathan
Department of Computer Science and Engineering
SCMS School of Engineering And Technology
Ernakulam, India
kavyasakb@gmail.com

Deepasree Varma P
Department of Computer Science and Engineering
SCMS School of Engineering and Technology
Ernakulam, India
deepasreevarma@gmail.com

***Abstract*— Plants are an inevitable part of ecosystem. Tulsi leaves play a major role in plant species. There exists different varieties of tulsi leaves mainly Rama tulsi, Krishna tulsi, Vana tulsi, Kapoor tulsi etc. This paper proposes a system for identifying and classifying different types of tulsi from its leaf images by analyzing color and texture. Rama tulsi, Krishna tulsi and ordinary tulsi are the three varieties of tulsi leaves used for classification. Initially, a dataset is formed from the images of leaves. The feature values of these images are computed and it becomes the training data for the system. When a test image is scanned its color and texture features are extracted. Using K nearest neighborhood algorithm the extracted features are compared with the training data to display category of the leaf. Texture features are obtained using Gray Level Co-occurrence matrix (GLCM) whereas color features are acquired using color moments. Fusions of these techniques are used in this paper.**

Keywords—Leaf Classification, KNN Classifier, Texture, GLCM, Color Moments

I. INTRODUCTION

Plant is one of the most important forms of life on earth. Plants maintain the balance of oxygen and carbon dioxide of earth's atmosphere. Plants can be classified according to the shapes, colors, textures and structures of their leaf, bark, flower, seedling and morph. Leaves contribute more diagnostic characteristics and hence it is used in most of the applications. Leaf identification contributes a major role in conservation of plants.

Tulsi leaves are grown in many parts of India. Tulsi leaves are used for various medicinal purposes like skin ailment, memory level increasing, digestion, acne and many. There are wide varieties of tulsi leaves like Krishna tulsi, Rama tulsi, Vana tulsi etc. Tulsi leaves can be identified based on its characteristics like color, shape, texture. Manual identification often consumes more time and is not accurate mostly. So there is an urgent need for an automated and reliable tool for identification and classification of tulsi leaves using available information. Till now, there is no existing system which classifies leaves of similar shape and type. This is the first project where the different types of leaves belonging to the same species are recognized by extracting color and texture features. This paper aims to identify and recognize tulsi leaves and classify them under their respective types.

Rashad et al. [1] proposed a novel approach for classifying plants by characterizing texture properties. A combined classifier learning vector quantization accompanied by radial basis function is used. The system recognizes and classifies a plant from a small section of the leaf. This system has an acceptance rate of 98.7% as compared to other methods. The problem with this system is that it doesn't consider noise.

Kadir et al. [2] introduced a technique which uses vein, shape, texture and color features for classification. Fourier descriptors of PFT and Geometric features are the shape features used in identification system. In this system given leaf image for classification is converted to gray level. Then, conversion from grayscale to binary form is done by using adaptive threshold. Several morphological operations are then performed to remove holes in the leaf caused by previous threasholding. AND operation is done between RGB image and binary image to obtain the leaf. Feature Extractor is used to extract the features contained in leaf. Probabilistic Neural Network (PNN) processes the features. An index representing plant is its result. The index is then translated into the name of the plant by the plant information getter. 93.75% accuracy is shown by the system.

Sumathi and Kumar [3] introduced a feature fusion technique with edge based feature extraction and Gabor filter in the frequency domain. The system trained extracted features using 10-fold cross validation. Radial Basis Function (RBF) and Classification and Regression Trees (CART) classifiers measured the accuracy. 85.93 % is the efficiency of RBF, has relatively less error for nine-class problem.

Beghin et al. [4] suggested an approach to combine easy methods which use texture and shape features. The shape-based method is used for calculating the dissimilarities between leaves from the extracted contour signature. Leaf shapes are analyzed from 2 contour signatures which used together provides sufficient amount of information about the leaf's shape. From the results of contour signatures, it is clear that leaves cannot be adequately classified based on shape alone. Texture is another main feature of the leaf. The macro-texture of the leaf is analyzed from the orientation of edge gradients. Incremental classification algorithm merged the results of these methods. It showed an accuracy of 81.1%.

A system for diagnosing medicinal plant leaves was designed by Arun et al. [5]. Grey textures, grey tone spatial dependency matrices (GTSDM) and Local Binary Pattern (LBP) operators were the features used for texture analysis. The plant leaves were classified based on feature values using six different classifiers. 94.7% efficiency was shown by the system when the features are merged.

This paper proposes a system where classification is done based on color and texture features. The leaf image is

978-1-5386-0576-9/18/$31.00 ©2018 IEEE

captured and is fed into the computer system. Values of color and texture features are calculated. The texture features are calculated from Gray Level Co-occurrence matrix (GLCM) of the image. These values are then matched with the values in the database. After comparison, the best match is displayed as result. All these methods are embedded into KNN software developed using Mat lab that covers the whole recognition process, from feature extraction to classification via KNN algorithm.

The rest of the paper is arranged into following sections: Section 2 provides a detailed description about proposed system and the techniques used. Experimental results are being discussed in section 3. Finally, Section 4 concludes this paper.

II. PROPOSED SYSTEM

The KNN software [17] aims at identifying and displaying the category of a particular tulsi leaf from a fed image. For this, the test image is compared with database images to find the image with the least difference in color and texture features. The main steps include the capturing leaf image, feature extraction, classification and displaying results.

A. Dataset collection and preprocessing

Image acquisition is the first step in leaf classification. Image acquisition includes collecting the digital color image of the leaf. The collection process is done by scanning 3 different varieties of tulsi leaves- Krishna tulsi, Rama tulsi, and ordinary tulsi. The digital color images are scanned to the computer in the jpeg format with dimension 375X725 resolutions to form the dataset. The aim of image pre-processing is to improve image data so that it can enhance the image features that are relevant for further processing. Here the leaves are scanned in the white background so as to avoid more pre-processing.

B. Feature Extraction

Tulsi leaves can be differentiated from its color. Color moments and Gray Level Co-occurrence matrix (GLCM) are the techniques used for feature extraction here.

Color moments characterize the color distribution in an image. Color moments are very effective for color-based analysis. Rama tulsi and Krishna tulsi have variations in the amount of green and red colors in their leaves. Color moments are computed in the same way as computing moments of a probability distribution. Most of the color distribution information is contained in low order moments. The first three color moments are usually used as features in image retrieval applications [6].

Mean: The average color in the image is considered as the first color moment [7]. It can be calculated by using the following formula:

$$Mean(\mu) = \sum_{j=1}^{N} \frac{1}{N} P_{ij} \tag{1}$$

where N is the number of pixels in the image and Pij is the value of the j-th pixel of the image at the i-th color channel.

Standard Deviation: The second color moment is the standard deviation. Standard deviation is the square root of the variance of the color distribution [6].

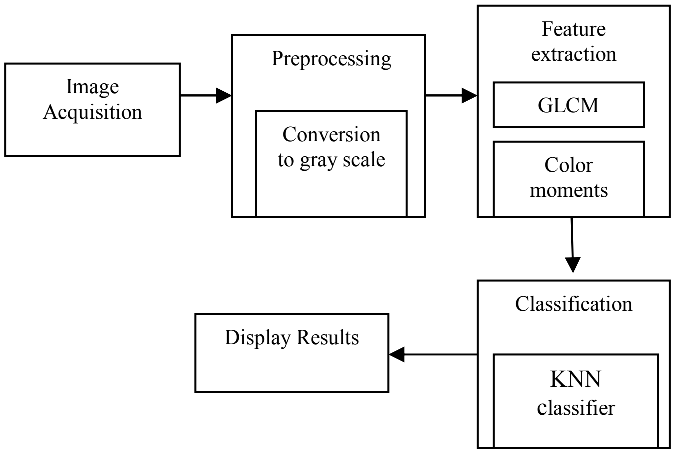

Fig. 1. Proposed system architecture

$$stddev(\sigma) = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(P_{ij} - \mu_i\right)^2\right)} \tag{2}$$

where μi is the first color moment for i-th color channel of the image.

HU invariant moments [7]: Seven moments proposed by HU are very useful for capturing the shape of the leaf. The moment of the image is the sum of moments of individual pixels. In the implementation, only the first four moments were used.

$$M(i,j) = \sum\sum(x^i)(y^j)I(x,y) \tag{3}$$

where (x, y) refer to co-ordinate of the leaf pixel. i, j can be varied to get different moments. I(x,y) is the intensity level at the pixel (x,y) of the image.

GLCM gives the properties of a texture but it is not useful for comparison of two textures. A number of texture features can be calculated from the GLCM. Some of the features tested were failed as the feature gives approximately same result for all the 3 types of tulsi leaves. 6 features used in this paper are Contrast, Correlation, Energy, Homogeneity, Entropy, PSNR VALUE.

$$\text{Contrast} = \sum_{i,j=0}^{N-1} P_{i,j}\,(i-j)^2 \tag{4}$$

$$\text{Correlation} = \sum_{i,j=0}^{N-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^{2})(\sigma_j^{2})}}\right] \tag{5}$$

$$\text{Energy} = \sum_{i,j=0} P(i,j)^2 \tag{6}$$

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} P(i,j)/R \tag{7}$$

where R is the range [0,1].

C. Classification

The classification approach used is, K-Nearest Neighbour (KNN) classifier based on the supervised classification. K Nearest Neighbour classifier calculates the minimum distance of a given point with other points to determine its class. For plant leaf classification, it calculates Euclidean distance between the test sample and training sample and display the category of training sample with shorter distance from test [8]. This is how it finds the class for test sample. The k-nearest neighbour algorithm is the simplest of all machine learning algorithms.

III. EXPERIMENTAL RESULTS

The proposed software was implemented and tested using real data and was found to be error free. Any user with the basic knowledge of using a personal computer or mobile can access it with ease. Also, it showed that the system is producing better results when the images are cropped.

Fig. 2. Sample dataset

The dataset contains a total of 60 images constituting 3 varieties of tulsi leaves. All the images are scanned in white background using a scanner so that the need for pre-processing can be avoided. The images are set to a standard resolution of 375x725 and are stored on 3 separate folders representing each type of Tulsi. Around 10% of the collected data are used for testing.

Initially, it classifies Rama tulsi and Krishna tulsi using color moments. Since the amount of green content is almost similar in both Rama tulsi and ordinary tulsi leaves, further classification can be done by analysing texture features using GLCM method. Grayscale of the image is generated for calculating GLCM. The texture features calculated are contrast, correlation, energy, homogeneity, entropy, and PSNR. The calculated GLCM values are normalized and are then loaded into a matrix. This process is repeated for all the images in the training set and the resulting values are loaded into a single matrix which becomes the database for KNN algorithm. Then using the same procedure it finds the texture feature values for a loaded test image. It will find and display the category of the leaf with the least difference after comparing these values with the ones in the database.

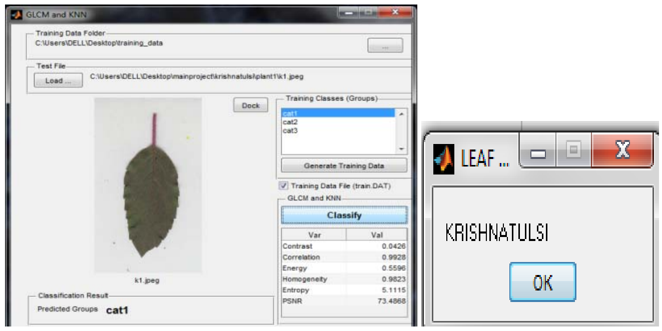

Fig. 3. Classification and results display

When the dataset is loaded, the KNN software generates 2 data files called group and train. Group file contains the category names of the leaves and train file contains the feature values generated. When a test image is loaded the software will compute the feature values for that particular leaf and are stored on to a file. Now a comparison is made between this file of values and the train data file and displays the category of the leaf with least difference using KNN algorithm.

A. Color Feature Extraction

The values obtained after using color moments function is shown below:

Values for Krishnatulsi leaves:

Table I COLOR VALUES FOR KRISHNA TULSI

Leaf	MeanR	StdR	MeanG	StdG	MeanB	StdB
Leaf1	84.778	5.9772	82.302	6.9830	65.842	6.024
Leaf2	88.6917	5.6345	82.113	6.4730	66.013	5.467
Leaf3	82.9402	6.5643	78.187	7.5195	62.634	6.336
Leaf4	85.8842	5.117	83.698	6.6648	67.635	5.112
Leaf5	88.8139	5.3889	81.766	5.9120	67.667	5.098

Values for Ramatulsi :

Table II COLOR VALUES FOR RAMA TULSI

Leaf	MeanR	StdR	MeanG	StdG	MeanB	StdB
Leaf1	95.778	6.372	108.252	6.9830	66.842	6.024
Leaf2	97.6917	7.5345	108.85	7.4730	67.013	7.467
Leaf3	97.9402	8.4643	109.59	8.5195	68.634	8.136
Leaf4	98.882	7.117	110.86	7.6648	68.635	6.112
Leaf5	98.8139	6.6889	110.94	6.9120	68.667	6.098

Mean value shows the average color in the image. From table, it is clear that green color distribution is more in Rama tulsi whereas it is low in case of Krishna tulsi.

Tables below shows the texture feature values obtained for all the three types of tulsi leaves.

B. Contrast

TABLE III. CONTRAST

Krishna tulsi	Rama tulsi	Ordinary tulsi
0.0398	0.1204	0.0380
0.0254	0.0670	0.0443
0.0371	0.0960	0.0348
0.0405	0.0903	0.0355
0.0346	0.0347	0.0267

Contrast gives the measure of local variations in an image. Contrast will be 0 for a constant image.

C. Correlation

TABLE IV. CORRELATION

Krishna tulsi	Rama tulsi	Ordinary tulsi
0.9929	0.9788	0.9931
0.9958	0.9845	0.9887
0.9943	0.9805	0.9918
0.9930	0.9807	0.9923
0.9944	0.9911	0.9938

It gives the measure of correlation of a pixel to its neighbor over the whole image.

D. Energy

TABLE V ENERGY

Krishna tulsi	Rama tulsi	Ordinary tulsi
0.5843	0.3811	0.5865
0.5246	0.5125	0.6986
0.4965	0.5141	0.6848
0.5466	0.5135	0.6595
0.5466	0.5252	0.6848

Energy when images are cropped:

TABLE VI. ENERGY WHEN CROPPED

Krishna tulsi	Rama tulsi	Ordinary tulsi
0.3665	0.276	0.205
0.457	0.2511	0.28
0.345	0.3230	0.388
0.434	0.266	0.344
0.456	0.3099	0.234

Energy detects the disorder in texture. Energy value decreases when images quality increases. From the table, it is clear that energy values decrease when the images are cropped.

E. Entropy

Entropy indicates the disorder or complexity of the image. Entropy is larger when the images are not textually uniform.

TABLE VII. ENTROPY

Krishna tulsi	Rama tulsi	Ordinary tulsi
5.1462	5.3324	4.8906
5.2475	4.7240	4.3498
5.3299	4.9968	4.5482
4.8310	5.0164	4.5469
4.9852	5.1679	4.3245

Entropy when images are cropped:

TABLE VIII. ENTROPY WHEN CROPPED

Krishna tulsi	Rama tulsi	Ordinary tulsi
6.03	5.83	6.25
5.73	6.105	5.92
6.05	5.44	6.33
4.85	5.88	5.65
4.08	5.45	5.76

From the table it is clear that the values of entropy are larger for cropped images as compared to that of original.

F. Homogenity

TABLE IX. HOMOGENEITY

Krishna tulsi	Rama tulsi	Ordinary tulsi
0.9837	0.9420	0.9831
0.9877	0.9687	0.9813
0.9837	0.9546	0.9865
0.9822	0.9576	0.9856
0.9847	0.9832	0.9888

Larger values of homogeneity show that the elements in images are similar. Homogeneity decreases if contrast increases while keeping energy as constant.

G. PSNR

PSNR values are greater for good quality images. Table shows the values of PSNR for different leaf category.

TABLE X. PSNR

Krishna tulsi	Rama tulsi	Ordinary tulsi
73.4836	73.4770	73.4825
73.4556	73.5136	73.5206
73.4389	73.5100	73.5127
73.5005	73.5196	73.5246
73.4767	73.5128	73.5042

IV. CONCLUSION AND FUTURE SCOPE

This paper proposed an automated system for plant identification using color and texture features extraction. This work aims at finding the exact match of the leaf from test parameters. The color and texture features are then extracted. These features are then compared with values of trained dataset to identify the category of tulsi leaf. GLCM technique is used here for feature extraction. 6 features of GLCM produced better results for classification. Color moments technique was used here for color feature extraction. The proposed system is effective with a recognition rate of 95%. The proposed approach also showed the difference in feature values when images are cropped to exact shape of the leaf. Hence it is clear that system shows better results when images are cropped.

Future works can be extended to all varieties of tulsi leaves. Also by using neural networks, the recognition rate can be increased. From the results, it is clear that the system shows better results when the images are cropped to exact leaf shape, so future works are to be done using cropped images with high quality.

ACKNOWLEDGMENT

We would like to thank the almighty for providing us this opportunity and granting us the strength to complete the work. Our deepest gratitude to the SSET management, teaching and non- teaching staff for providing us with the resources that enabled us to complete the work within specified time. We are very grateful to all our well wishers who supported us directly and indirectly during the work.

REFERENCES

[1] M. Z. Rashad, B. S. el-Desouky,and Manal S. Khawasik, "Plants images classification based on textural features using combined classifier", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No. 4, August 2011,pp.93-100.

[2] Abdul Kadir, Lukito Edi Nugroho, and Paulus Insap Santosa, "Leaf classification using shape, color, and texture", International Journal of Computer Trends & Technology (IJCTT), July-August 2011,pp.225-230

[3] C. S. Sumathi and A. V. Senthil Kumar, "Edge and texture fusion for plant leaf classification", International Journal of Computer Science and Telecommunications, Vol 3, Issue 6, June 2012,pp. 6-9.

[4] T. Beghin, J. S. Cope, S. Barman, & P. Remagnino, "Shape and texture based plant leaf classification", Advanced Concepts for Intelligent Vision Systems (ACVIS),Vol 6475,2010,pp.45-353.

[5] C. H. Arun, W. R. Sam Emmanuel, and D. Christopher Durairaj: "Texture feature extraction for identification of medicinal plants and comparison of different classifiers", International Journal of Computer Application ,Vol 62, Jan 2012

[6] Hong-Jiang Zhang Jufu Feng Hui Yu, Mingjing Li. "Color texture moments for content-based image retrieval". pages 929– 932, September 2002.

[7] W. Y. Ma, H. J. Zhang, "Content-based Image Indexing and Retrieval": In Handbook of Multimedia Computing, Borko Furht. Ed., CRC Press, 1998

[8] Prof. Meeta Kumar, Mrunali Kamble, Shubhada Pawar, Prajakta Patil, Neha Bonde, "Survey on techniques for plant leaf classification", International Journal of Modern Engineering Research (IJMER), Vol 1, Issue 2,pp-538-544.

[9] H.Fu and Z. Chi, "Combined thresholding and neural network approach for vein pattern extraction from leaf images", IEEE Proc-Vis. Image Signal Process., Vol.153, No 6, December 2006, pp.881-892.

[10] B. S. Manjunath, W. Y. Ma, "Texture features for browsing and retrieval of image data", IEEE Trans on PAMI, Vol. 18, No. 8, pp. 837-842, 1996

[11] Y. Ye, C. Chen, C.-T. Li, H. Fu, and Z. Chi, "A computerized plant species recognition system," In proceedings of the 2004 International Symposium on Intelligent Multimedia, Video, and Speech Processing, Hong Kong, October, 2004

[12] X.-F. Wang, J.-X. Du, and G.-J. Zhang,"Recognition of leaf images based on shape features using a hyper sphere classifier," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS 3644. Springer, 2005

[13] R. S. Choras,"Image feature extraction techniques and their application for CBIR and biometrics systems", International Journal of Biology and Biomedical Engineering , vol. 1 (1), pp. 6-16, 2007.

[14] Gu, X., Du, J.X., Wang, X.F."Leaf recognition based on the combination of wavelet transform and gaussian interpolation". Advances In Intelligent Computing 3644,253-262 ,2005.

[15] Arivazhagan, S., Ganesan, L., Priyal, S.P. "Texture classification using Gabor wavelets based rotation invariant features". Pattern Recognition Letters 27, 1976– 1982 (2006).

[16] Q. Wu, C. Zhou, & C. Wang., "Feature extraction and automatic recognition of plant leaf using artificial neural network", Avances en Ciencias de la Computacion, pp. 5-12, 200.

[17] http://indogateways.com/matlab/?mc=a71bb845451d9e60b9baeff71203fad736cbe1