IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-2, Feb 2019 E-ISSN: 2347-2693

Plant Leaf Analysis Based on Color Histogram and Cooccurrence Matrices

Renuka R. Londhe

Dept. of Computer Science and IT, Rajarshi Shahu Mahavidyalaya (Autonomous), Latur, 413512 India

Corresponding Author: renu_sanskruti29@rediffmail.com, Tel.: +91-9545226333

DOI: https://doi.org/10.26438/ijcse/v7i2.153157 | Available online at: www.ijcseonline.org

Accepted: 24/Feb/2019, Published: 28/Feb/2019

***Abstract*:** Automatic identification of plant is very useful for environmentalists, natural scientists, biologists, food engineers, amateur botanists, educators and doctors (*Ayurvedacharya*). In this paper a computer based application was developed to automatically identify herbal plant type by the photographs of plant leaves. The leaf image used for analysis can be either a database digital image or photograph recorded by camera. The image used was of single leaf with light and white background. The leaf image analysis has been performed with MATLAB 2016. The procedure comprised of analysis of leaf image segmentation, feature extraction from Shape, Color histogram and Cooccurrence Matrices respectively. Shape of leaf is the furthermost widespread feature used in identification. The Leaf analysis has been performed for 25 herbal medicinal plant leaves from Folio database. The Shape feature was extracted using edge detection operator Sobel; and to record the color statistical features, color histogram and co-occurrence matrices with statistical parameters. The results of this article will be useful to identify the leaves of different types of plants.

Keywords— Leaf Analysis, Color Cooccurrence Matrix, Color Histogram, Statistical Features, Shape Features

I. INTRODUCTION

Plants are crucial part of the environment. A large number of plants exist in environment. The herbal and medicinal plants play a crucial role in *Ayurvedic* medicine. Most of the herbal medicinal plants are on the verge on destroying due to the environmental pollution, biodiversity loss, global warming, effects of fast urban development, and various environmental damages. Therefore, there is urgent need for protect, identity the features and useful properties of these plants. The advanced computer vision techniques can be applied to obtain the botanical information of the rare plants including plant classification, numerous features of the plant and make this information open and convenient to researchers, farmers, botanists, students and ayurvedic doctors [1]. Thus the plant analysis and identification prime task for researchers.

There are various plant parts including seeds, flowers, leaves, and fruits which are used for plant identification. In the preset research the plant leaves are designated to acquire the plant features. The leaves can be obtained at ease, scanned and also it comprises of more exciting information useful for plant classification. The leaf images are directed to the Matlab software and then through image processing tools, leaf features are extracted for identification of the plants.

The paper is organized as follows, Section I is introductory in nature, Section II comprises the work related to plant recognition system, Section III contain the some measures of database and feature extraction methods such as color histogram and co-occurrence matrices as well as the architecture and essential steps of leaf analysis, section IV describes results and discussion., Section V deals with conclusions based on the current work.

II. RELATED WORK

Numerous methods have been used for plant identification. Most of the methods use leaf for extracting features. Few groups have used shape description method, while others have used texture features, geometrical features, color features, vein structures, etc [2]. Wang and colleagues [3] have introduced the shape descriptor method for online identification of the plant leaf on mobile platform. Wu and coworkers [4] have extracted 12 leaf features and orthogonised them into 5 principal variables through Principal Component Analysis (PCA) and Probabilistic Neural Network (PNN) as a classifier.

Chaki et al. [5] used Gabor Filter for recognizing the plant types. The Convolving Gabor filters with leaf images, real and imaginary parts of resultant signal are acquired.

Complete variance between leaves images used as a feature. The shape of Leaf is subject to distortions produced by disease, insects or even human and mechanical damage. Consequently, color and texture features are considered to progress leaf-based plant identification.

Color is the apparent morphological feature of a leaf. In color moments in Red, Green and Blue (RGB) comprises mean, standard deviation, skewness and kurtosis are implemented to describe color features of plant leaves [6, 7], making it appropriate for real-time applications. Texture is additional feature that can be applied in plant identification to designate the vein structure or leaf's surface. Analogous to colour feature, texture is measured as an added feature to better illustrate the leaf properties [8].

Kadir and coworkers [9] extracted contrast, inverse different moment, angular second moment, entropy and correlation from the gray-level cooccurrence matrix of the leaf. The latest research have exploited Haar wavelet, Radial Basis function [10], Gabor filters, and two dimensional multifractal detrended fluctuation analysis [11] to progress identification correctness. Lin et al. [12] used various image processing techniques for leaf disease detection. They developed Android application, client-server architecture and used sensor technology for application. Thomous and colleagues [13] discussed various edge detection techniques on an image of a leaf to study the venation pattern of that leaf. Detection of the veins have been compared and analyzed to find the efficiency and accuracy of edge detection techniques. They found that the step edges and ridge edges by taking the Gaussian and the first and second order derivative of the Gaussian of the image.

In this paper leaf identification algorithm has been employed by extracting the leaf features from shape and color of leaves. Edge detection operator is applied on each leaf and number of pixels denotes the edge is computed. To extract the texture based feature cooccurrence matrix is used for the surface of image and some features are extracted using color histogram.

III. METHODOLOGY

Database: In this work we have used Folio database. The leaves of 25 types of plants are captured from camera in natural sunlight with resolution of 4128×2322. The captured images are from the herbal and medicinal plants such as Beaumier du perou, Eggplant, Fruit citere, Guava, Hibiscus, Betel, Rose and so on. From the plane background image, area of the leaf is extracted and stored for further processing.

Feature Extraction: To extract the features from processed images, edge detection, color histogram and cooccurrence matrices are used. As we know that the shape of leaf plays a significant role in leaf identification. To extract the shape of image, we have applied Sobel edge detection operator on the images and shapes of images are observed. Through the program we find out the number of pixels denoting the edges of leaf and we have received the different values for each and every leaf and these are considered as one of the dominant features for leaf identification.

Edge Detection:

Edges are noteworthy to local modifications of intensity in the image. Edges typically occur on the boundary between two different regions in an image. Edge detection is the approach used for segmenting images based on local changes in the intensity.

Edge detection produces a line drawing of a scene from an image of that scene. Using edge detection important features can be extracted from the edges of an image for example corners, lines and curves. These features are used by higher-level computer vision algorithms such as recognition.

There are different techniques for edge detection. We have used Sobel edge detection operator. The operator consists of a pair of 3×3 convolution kernels as shown in figure 1. Figure 2 illustrate the image after applying edge detection operator.

-1	**0**	**+1**		**+1**	**+2**	**+1**
-2	**0**	**+2**		**0**	**0**	**0**
-1	**0**	**+1**		**-1**	**-2**	**-1**

Gx Gy

Figure 1 Masks used by Sobel operator

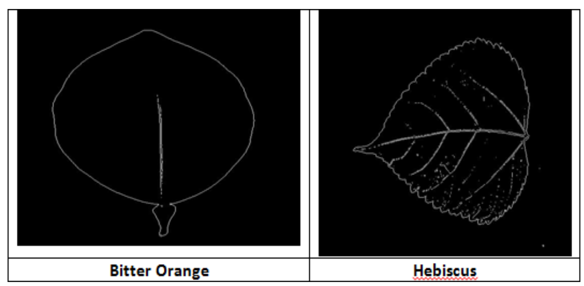

Figure 2 Sample images of edges or shapes of leaves

Color Histogram: The histogram of a digital image with L total possible intensity levels in the range [0, G] is defined as the discrete function

$$\mathbf{h(r_k) = n_k}$$

Where r_k is the k^{th} intensity level in the interval [0, G] and n_k is the number of pixels in the image whose intensity level is r_k.

Normalized histograms are obtained simply by dividing all elements of $h(r_k)$ by the total number of pixels in the image, which we denote by n:

$$P(r_k) = \frac{h(r_k)}{n} \quad (1)$$
$$= \frac{n_k}{n}$$
for k = 1,2,…L.

In colour histogram we need to extract red, green, blue components of an images and then we have to compute the histogram for all the components as per above procedure. Color histogram is obtained for every leaf and dominant features are collected from red, green and blue histograms. Total seven features are collected from each type of leaf using histogram. Amongst them one denotes the number of absent intensity levels and remaining six are the number of pixels of two bins from red green and blue histograms respectively.

Color Co-occurrence Matrix: One of the simplest methods for defining texture is to utilise statistical moments of the intensity histogram of an image. The use of histograms in calculation results into measurement of texture carrying the information about intensity distribution only, and does not contain the information about relative position of pixels with respect to each other. The statistical methods like co-occurrence matrix provide useful information about the relative position of the adjacent pixels in an image.

Given an image f, of size N×N, the co-occurrence, matrix P can be defined as

$$P(x,y)=\sum_{x=1}^{N}\sum_{y=1}^{N}\begin{cases}1, iff\ (x+\Delta_x, y+\Delta_y)=j\\ 0 otherwise\end{cases} \tag{2}$$

The following figure 3 shows an example of construction of a cooccurrence matrix using L=8 and a position operator Q defined as one pixel instantly to the right. We see that element (1, 1) in the GLCM contains the value 1 because there is only one instance in the image where two, horizontally adjacent pixels have the values 1 and 1. Element (1, 2) in the GLCM contains the value 2 because there are two instances in the image where two, horizontally adjacent pixels have the values 1 and 2 and so on.

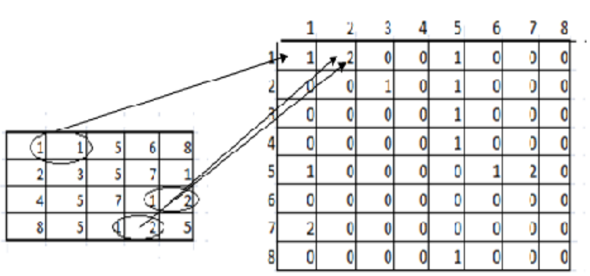

Figure 3 Construction of co-occurrence Matrix

In this experiment red, green and blue channels are separated from the RGB image and co-occurrence matrices are extracted for each channel. Figure 4 shows equivalent co-occurrence matrix of the red channel sample image.

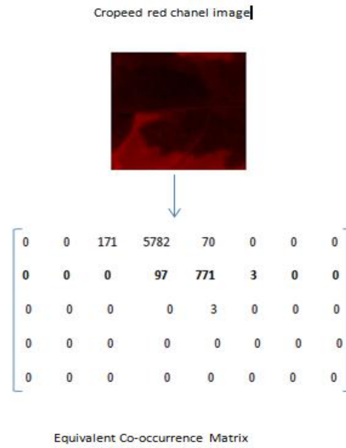

Figure 4. Cropped image and cooccurrence Matrix.

From these co-occurrence matrices four Haralick features including contrast correlation energy and homogeneity are calculated.

Contrast: Returns a measure of the intensity contrast amongst a pixel and its neighbor over the whole image.

Range = [0 (size (GLCM, 1)-1)^2] For a constant image the contrast is 0 (zero). The property contrast is also recognized as variance and inertia.

$$\sum_{i,j}|i-j|2p(i,j)$$

Correlation: It is a measure of correlation of a pixel with its neighbor over the complete image. Range = [-1 1]
For a perfectly positively or negatively correlated image the Correlation is 1 or -1 respectively.
For a constant image the Correlation is NaN .

$$\sum_{i,j}\frac{(i-\mu i)(j-\mu j)p(i,j)}{\sigma i\sigma j}$$

Energy: Returns the sum of squared elements in the GLCM. Range = [0 1] For a constant image, the Energy is 1.

$$\sum_{i\ j}p(i\ j)2$$

The property Energy is also identified as *uniformity of energy*, *uniformity* and *angular second moment*.

Homogeneity: It is a measure the closeness of the distribution of elements in the GLCM to the GLCM diagonal. Range = [0 1].

$$\sum_{i\ j}\frac{p(i,j)}{1+|i-j|}$$

Where p is the number of Grey level co-occurrence matrices. This feature extraction process is shown in Figure 5.

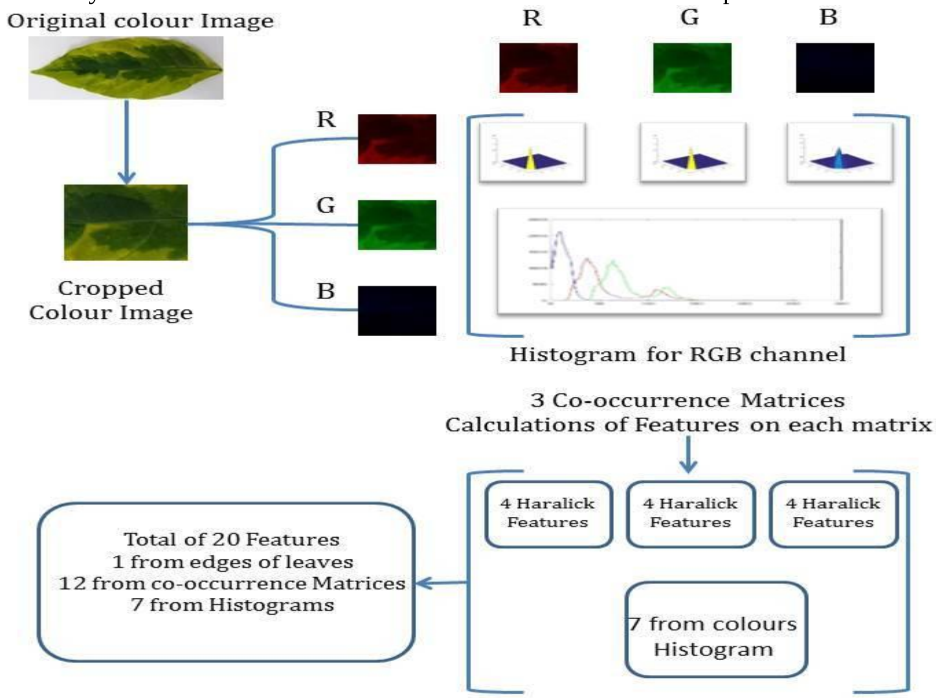

Figure 5: Steps for the process of Feature Extraction

Table 1: The feature vectors of twenty five leaves with 8 features.

	Shape	AL	RB	GB	BB	RB	GB	BB
Leaf1	1970	169	1	0	318	131	1441	1
Leaf2	1431	124	2227	23	92	10	13	0
Leaf3	1791	141	38	0	636	86	310	1
Leaf4	1755	210	2495	0	0	0	6	0
Leaf5	1312	175	2208	2	65	6	54	0
Leaf6	1025	107	1042	321	0	80	130	0
Leaf7	817	74	644	59	369	87	155	96
Leaf8	846	172	460	1107	1267	206	34	2
Leaf9	574	142	3165	559	243	7	21	1
Leaf10	879	127	493	193	1144	24	254	0
Leaf11	1879	212	2356	0	1	0	15	0
Leaf12	1615	72	482	736	40	61	109	0
Leaf13	1369	172	0	0	1476	0	0	0
Leaf14	1067	184	0	0	1931	19	179	0
Leaf15	991	141	985	0	854	94	570	42
Leaf16	663	102	564	219	40	13	40	4
Leaf17	603	152	442	26	70	71	1021	0
Leaf18	802	163	1333	721	105	21	145	0
Leaf19	1034	145	512	494	359	50	147	12
Leaf20	929	101	0	0	70	62	150	68
Leaf21	634	174	351	0	147	52	244	0
Leaf22	580	205	1083	3	1	0	0	0
Leaf23	513	162	915	0	291	66	446	2
Leaf24	687	87	2945	1	37	4	35	2
Leaf25	730	194	35	0	3	2	246	0

AL; Absent Levels, RB; Red Bin, GB; Green Bin, BB; Blue Bin

Table 2: The feature vectors of twenty five leaves with 12 features.

	CR	COR	ER	HR	CG	COG	EG	HG	CB	COB	EB	HB
Leaf1	0.15	0.68	0.46	0.93	0.16	0.73	0.35	0.92	0.16	0.73	0.35	0.92
Leaf2	0.05	0.79	0.86	0.98	0.12	0.82	0.41	0.94	0.12	0.82	0.41	0.94
Leaf3	0.19	0.66	0.34	0.91	0.14	0.58	0.59	0.93	0.14	0.58	0.59	0.93
Leaf4	0.00	0.44	0.99	1.00	0.04	0.65	0.83	0.98	0.04	0.65	0.83	0.98
Leaf5	0.02	0.83	0.89	0.99	0.08	0.86	0.42	0.96	0.08	0.86	0.42	0.96
Leaf6	0.05	0.97	0.37	0.97	0.07	0.95	0.30	0.97	0.07	0.95	0.30	0.97
Leaf7	0.20	0.89	0.29	0.90	0.18	0.90	0.27	0.91	0.18	0.90	0.27	0.91
Leaf8	0.08	0.86	0.41	0.96	0.06	0.85	0.57	0.97	0.06	0.85	0.57	0.97
Leaf9	0.04	0.70	0.84	0.98	0.05	0.64	0.84	0.97	0.05	0.64	0.84	0.97
Leaf10	0.16	0.82	0.32	0.92	0.14	0.84	0.38	0.93	0.14	0.84	0.38	0.93
Leaf11	0.01	0.63	0.95	0.99	0.01	0.85	0.93	1.00	0.01	0.85	0.93	1.00
Leaf12	0.13	0.95	0.31	0.93	0.09	0.95	0.48	0.95	0.09	0.95	0.48	0.95
Leaf13	0.03	0.90	0.75	0.99	0.03	0.84	0.80	0.99	0.03	0.84	0.80	0.99
Leaf14	0.09	0.80	0.45	0.95	0.02	0.65	0.93	0.99	0.02	0.65	0.93	0.99
Leaf15	0.15	0.78	0.41	0.93	0.14	0.79	0.45	0.93	0.14	0.79	0.45	0.93
Leaf16	0.12	0.86	0.39	0.94	0.11	0.83	0.59	0.95	0.11	0.83	0.59	0.95
Leaf17	0.08	0.86	0.41	0.96	0.10	0.85	0.44	0.95	0.10	0.85	0.44	0.95
Leaf18	0.07	0.79	0.63	0.96	0.08	0.85	0.44	0.96	0.08	0.85	0.44	0.96
Leaf19	0.13	0.85	0.36	0.94	0.05	0.93	0.46	0.98	0.05	0.93	0.46	0.98
Leaf20	0.06	0.93	0.53	0.97	0.03	0.95	0.80	0.98	0.03	0.95	0.80	0.98
Leaf21	0.02	0.94	0.66	0.99	0.01	0.95	0.77	0.99	0.01	0.95	0.77	0.99
Leaf22	0.07	0.80	0.57	0.96	0.05	0.87	0.54	0.97	0.05	0.87	0.54	0.97
Leaf23	0.08	0.78	0.58	0.96	0.07	0.68	0.72	0.96	0.07	0.68	0.72	0.96
Leaf24	0.04	0.82	0.86	0.98	0.08	0.78	0.62	0.96	0.08	0.78	0.62	0.96
Leaf25	0.13	0.74	0.40	0.94	0.03	0.71	0.88	0.99	0.03	0.71	0.88	0.99

CR; Contrast for Red, COR; Correlation for Red, ER; Energy for Red, HR; Homogeneity for Red, CG; Contrast for Green, COG; Correlation for Green, EG; Energy for Green, HG; Homogeneity for Green, CB; Contrast for Blue, COB; Correlation for Blue, EB; Energy for Blue, HB; Homogeneity for Blue,

IV. RESULTS AND DISCUSSION

From each leaf a total of 20 features are extracted, first feature represents the shape of the leaf through pixels computed from the edges of images. The second feature designates total number of absent intensity levels. The features three to eight designates parameters of two bins from red, green and blue images. Table 1 shows the feature vectors of twenty five leaves with first 8 features. The Features nine to twenty gives the four Haralick features from red, green and blue images. The features discussed in Table 2 are the contrast, correlation, energy and Homogeneity. Table 2 shows the feature vectors of twenty five leaves with 9 to 20 features. It is observed that the first column of feature's table have each value as different one, it indicates the shape feature is dominant. When the two histograms of different images are compared, it gives the difference in number of pixels for the different levels.

V. CONCLUSIONS AND FUTURE SCOPE

From above study it can be concluded that the features such as shape, texture and color are the key features for automatic identification of plant leaf. We have successfully applied MATLAB16 for extraction of features of some medicinal herbal plants. A simple table of feature vectors is extracted for the analysis. The Harlick features obtained from co-occurrence matrices are different for every leaf. Feature extraction method may be the useful parameter for any plant identification system. The proposed methodology should also be conducted on any of the publicly available image database.

REFERENCES

[1] Mzoughi, I. Yahiaoui, N. Boujemaa and E. Zagrouba, *"Advanced tree species identification using multiple leaf parts image queries",* IEEE ICIP 2013.

[2] Bhardwaj, M. kaur, *"A review on plant recognition and classification techniques using leaf images",* International Journal of Engineering Trends and Technology, Vol. 4, Issue 2, 2013.

[3] Wang, D. Brown, Y. Gao and J. L. Salle, *"Mobile plant leaf identification using smart-phones"*, IEEE ICIP 2013.

[4] S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang, and Q. L. Xiang, *"A leaf recognition algorithm for plant classification using probabilistic neural network",* IEEE ISSPIT, pp. 11– 16, 2007.

[5] J. Chaki, and R. Parekh, *"Plant leaf recognition using Gabor filter",* International Journal of Computer Applications Vol. 56 issue10.

[6] Ghasab MAJ, Khamis S, Mohammad F, Fariman HJ, "*Feature decision-making ant colony optimization system for an automated recognition of plant species*", Expert Systems with Applications , Vol. 42, issue 5, pp 2361-2370, 2015.

[7] Stricker MA, Orengo M., *"Similarity of color images", IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology*, pp 381–392, 1995.

[8] Chaki J, Parekh R, Bhattacharya R., *"Plant leaf recognition using texture and shape features with neural classifiers",* Pattern Recognition Letters, Vol 58, pp 61-68, 2015.

[9] Kadir A, Nugroho LE, Susanto A, Santosa PI, *"Experiments of Zernike moments for leaf identification",* Journal of Theoretical and Applied Information Technology, Vol 41, issue 1, pp 83-93, 2012.

[10] Rashad MZ, el-Desouky BS. Khawasik MS, *"Plants images classification based on textural features using combined classifier",* International Journal of Computer Science & Information Technology , Vol. 3, issue 4, pp 93-100, 2011.

[11] Wang F, Liao DW, Li JW, Liao GP, "*Two-dimensional multifractal detrended fluctuation analysis for plant identification",* Plant Methods , Vol. 11, issue 1, pp 12-18, 2015.

[12] Asha Patil, Kalyani Patil, Kalpesh Lad, "*Leaf Disease detection using Image Processing Techniques*", International Journal of Scientific Research in Computer Science and Engineering, Vol.06, Issue.01, pp.33-36, 2018.

[13] C.T. Lin, M. Thomous, "*Study and Overview of Venation of leaf using Image Processing*", International Journal of Scientific Research in Computer Science and Engineering, Vol.4, Issue.5, pp.25-30, 2016.

Authors Profile

Renuka R. Londhe pursed her Bachelor of Science and Master of Science from SRTMU, University, Nanded, India, in year 2002 and 2004 respectively. She is also pursed M. Phil. and Ph. D. in the year 2009 and 2013 respectively. She is currently working as Assistant Professor of Computer Science in Department of Computer Science and IT, Rajarshi Shahu Mahavidyalaya (Autonomous), Latur. She has published more than 10 research papers in reputed international journals and conference proceedings. Her main research work focuses on Digital Image Processing. She has 13 years of teaching experience and 5 years of Research Experience.

.

