Classification of Selected Medicinal Plants Leaf Using Image Processing

A Gopal [1]
[1] CSIR - Central Electronics Engineering Research Institute
CSIR-CEERI Chennai Centre, CSIR Madras Complex
Chennai – 600 113, India
e-mail:agopal@hotmail.com

S. Prudhveeswar Reddy [2], V. Gayatri [2]
[2] Student Trainee, CSIR-CEERI Chennai Centre,
CSIR Madras Complex,
Chennai – 600 113, India

Abstract: **Plants are an indispensable part of our ecosystem and the dwindling number of plant varieties is a serious concern. To conserve plants, their rapid identification by botanists is a must, thus a tool is needed which could identify plants using easily available information. There is a growing scientific consensus that plant habitats have been altered and species are disappearing at rates never witnessed before. The biodiversity crisis is not just about the perilous state of plant species but also of the specialists who know them This initially requires data about various plant varieties, so that they could be monitored, protected and can be used for future. Plants form the backbone of Ayurveda and today's Modern day medicine and are a great source of revenue. Due to Deforestation and Pollution, lot of medicinal plant leaves have almost become extinct. So, there is an urgent need for us to identify them and regrow them for the use of future generations. Leaf Identification by mechanical means often leads to wrong identification. Due to growing illegal trade and malpractices in the crude drug industry on one hand and lack of sufficient experts on the other hand, an automated and reliable identification and classification mechanism in order to handle the bulk of data and to curb the malpractices is needed. The following paper aims at implementing such system using image processing with images of the plant leaves as a basis of classification. The software returns the closest match to the query. The proposed algorithm is implemented and the efficiency of the system is found by testing it on 10 different plant species. The software is trained with 100 (10 number of each plant species) leaves and tested with 50 (tested with different plant species) leaves. The efficiency of the implementation of the proposed algorithms is found to be 92%.**

Keywords: **Leaf classification, colour features, shape features, image processing, dissimilarity measures.**

I. INTRODUCTION

Leaf is an important parameter while studying plant nutrition, plant competition, plant soil-water relations, plant protection measures, crop ecosystems, respiration rate, transpiration rate and photosynthesis. To conserve plant species, their identification is the first step. Therefore, it is highly essential to have an object recognition system to identify various species and protect them from being endangered. The developments in the field of computer vision can be utilized for this purpose. Computer vision made duplication of human vision possible by electronically capturing an image. The importance of leaf area determination in plant sciences has stimulated many methods for leaf area measurement like grid counting [1], blue printing, photographing, length and width correlation [1,2], and usage of electronic devices etc. In general, computer based systems consists of a scanner or digital camera used to capture the image of the leaf, which is analyzed by the computer using specifically developed software [3]. Use of statistical discriminant analysis along with color based clustering and neural networks like those used in [4] for classification of a flowered plant and a cactus plant. Curvature Scale Space (CSS) technique and k-NN classifiers [5], were used to classify chrysanthemum leaves [6]. Using a shape characterization function called centroid-contour distance curve and the object eccentricity for leaf image retrieval, like in [7]. Incorporating HSV color space and SIFT [8] to extract various features of the leaf image, like in [9]. . Features that could be used for differentiation between leaves are: shape, edge, tip, base, surface, trichomes and arrangement on stem divisions of the main blade, color, texture and venation. Currently known are 52 different shapes of the leaves, 13 types of edges, 9 types of tips, 11 types of bases,12 types of surfaces and 28 types of trichomes. [10] Further, arrangement on stem can be of 4 types-Alternate, Opposite, Whorled and used in identification of a leaf and has 9 types. [10] The proposed method uses the combination of shape & color features to classify the query leaf on comparison with the leaves present in the database.

II. IDENTIFICATION METHODS

There are several methods by which plant leaves and other materials can be identified. Some of the most popular methods among these are: Spectroscopy, Chemical Identification and Optical Identification. Chemical Identification methods are divided into classical methods and instrumental methods. The classical methods are further divided into Chemical tests and flame tests. For the identification of plants, Auger spectroscopy and Hyper-spectral Imaging can be used. Our proposed method uses optical methods. Digital image processing comprises of using computer algorithms to process digital images. It is superior to analog image processing as it allows a wider range of algorithms to be applied to the input data and can avoid signal distortion and build-up of noise during processing. Using optical methods, several steps are performed such as Classification, Feature extraction and Pattern recognition.

III. IMPORTANT PARAMETERS/FEATURES OF LEAF

A. Shape

978-1-4673-2322-2/12/$31.00 ©2012 IEEE

This is one of the best features to identify a leaf. Using highly developed algorithms to identify the great variety of shape is possible through image processing. This will also encompass the edge, tip and base features of the leaf.

B. *Venation*

This varies a lot in leaves. Although belonging to one of the nine categories of venation, this property has slight differences that can pinpoint to a particular leaf.

C. *Texture*

Texture is a powerful regional descriptor that helps in retrieval process. Texture on its own does not have the capability of finding similar images, but it can be used to differentiate textured images from non-textured images and then combined with another visual attribute like colour to make the retrieval more effective.

D. *Colour*

Colour is a good identification parameter when there is a variation of colour in leaves of different species. But ,Care must be taken to use colour features only with fresh leaves because as leaves age, they may change the colour due to loss of chlorophyll and the presence of accessory pigments.

IV. PARAMETERS USED IN TH IMPLEMENTATION OF PROPOSED SOFTWARE

A. *Shape features*

1) *Aspect ratio*

It is the ratio of width of the leaf to the length of the leaf. In other words, it can be thought off as ratio of maximum axial length to minimum axial length of the leaf. It is also referred to as Slimness.

Aspect ratio =Maximum axial length/Minimum axial length

2) *Compactness*

It is defined as the ratio of the product of area with 4π to the square of perimeter .It is also referred to as roundness

$$Compactness(roundness) = \frac{4\pi * Area}{(perimeter)^2}; \quad (1)$$

3) *Dispersion*

It is defined as the ratio between the radius of the maximum circle enclosing the region and minimum circle that can be contained in the region. This measure increases as region spreads. However, Dispersion is insensitive to slight discontinuity in the shape such as crack in the leaf.

$$Dispersion = \frac{\max\left(\sqrt{(xi-x')^2+(yi-y')^2}\right)}{\min\left(\sqrt{(xi-x')^2+(yi-y')^2}\right)}; \quad (2)$$

Where (xi, yi) is the co-ordinate of the pixel in the leaf and (x', y') is the centroid of the leaf.

4) *Centroid*

The centroid co-ordinates of the figure are obtained and labelled as centroidx and centroidy

5) *Eccentricity*

Eccentricity is a characteristic feature of any conic section .It is given by the formula

$$Eccentricity = \sqrt{\left(1-\left(\frac{b}{a}\right)^2\right)}; \quad (3)$$

Where 'b' and' a 'refer to minimum and maximum axial length respectively for an ellipse. Similar formulas can be derived for other conic sections.

6) *HU invariant moments*

Seven moments proposed by Hu [11] are very useful to capture shape of the leaf .The moments are invariant to rotation, translation and scaling. The moment of the image is the sum of moments of individual pixels. In the implementation, only first four moments were used.

$$M(i,j) = \sum\sum(x^i)(y^j)I(x,y); \quad (4)$$

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

The first four moments used were

$$\mu1 = M(2,0) + M(0,2); \quad (5)$$

$$\mu2 = (M(2,0) - M(0,2))^2 + (2M(1,1))^2; \quad (6)$$

$$\mu3 = (M(3,0) - 3*M(1,2))^2 + (M(0,3) - 3*M(2,1))^2; \quad (7)$$

$$\mu4 = (M(3,0) + M(1,2))^2 + (M(0,3) + M(2,1))^2; \quad (8)$$

The mean of the first four moments is taken as a parameter for implementation of the algorithm

7) *Polar fourier transform*

Polar Fourier Transform (PFT) is very useful to capture shape of the leaf. The descriptors extracted from PFT are invariant under the actions of translation, scaling and rotation. PFT used in the implementation is defined as follows

$$PF2(\rho,\phi) = \sum\sum f(\rho,\phi)\exp\left(j2\pi\left(\frac{r}{R}\rho + \frac{2\pi}{T}\phi\right)\right); \quad (9)$$

Where R is radial frequency resolution and T is angular frequency resolution.

Computation of PFT is as follows: For example, there is an image I={f(x,y); 0≤x<M, 0≤y<N}.Firstly the image is converted from Cartesian space to polar space I'={f(r,θ); 0≤ r <R, 0≤θ<2π} where R is maximum radius from the centre of the sphere. The origin of polar space becomes as centre of space to get translation invariant. Rotation invariance can be achieved by ignoring the phase coefficient. Consequently only the magnitudes of coefficients are retained. Meanwhile to get scale invariance, the first magnitude value is normalised by the area of the circle and all magnitudes are normalised by magnitude of first coefficient. So, the Fourier descriptors are

FD={PF(0,0)/(2*PI*R^2),PF(0,1)/PF(0,0),.........,PF(m,n)/PF (0,0)}

Where 'm'is the maximum number of radial frequencies and 'n' is the number of angular frequencies.

B. *Colour features*

Colour moments represent colour features to characterise a colour image. Features that can be involved are Mean, Standard deviation, skewness and kurtosis. For RGB colour space, the features can be extracted from each plane Red, Green and Blue. The formulas to capture those moments are

$$Mean(\mu) = \frac{\sum\sum P(i,j)}{(MN)}; \quad (10)$$

$$Stddev(\sigma) = \sqrt{\left(\frac{1}{MN}\sum\sum(P(i,j)-\mu)^2\right)}; \quad (11)$$

$$Skewness(\theta) = \frac{\sum\sum\left((P(i,j)-\mu)^3\right)}{\left(MN\sigma^3\right)}; \quad (12)$$

$$Kurtosis(\gamma) = \frac{\sum\sum\left((P(i,j)-\mu)^4\right)}{MN\sigma^4}; \quad (13)$$

Where M and N are the dimension of the image.P(i,j) is the values of the colour on column i and row j. In the implementation the difference between green mean and red mean, difference between green mean and blue mean, difference between green standard deviation and red standard deviation, difference between green standard deviation and blue standard deviation, skewness and kurtosis of the three colour planes are used as parameters in the implementation of the software.

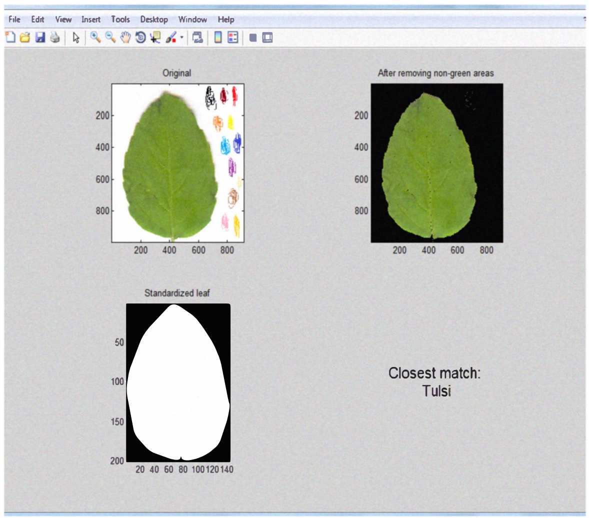
V. PROPOSED ALGORITHM

The aim of the software is to identify and display which plant a particular leaf belongs to from a fed image. This can be fulfilled by following some abstract steps, each with some sub-steps, which together make up the algorithm of the software. The abstract steps include capturing of leaf image, feature extraction, pattern recognition, classification and displaying results.

First the image of the leaf is captured using a digital scanner. Pre-processing, to remove all the excess information from the image in order to reduce the complexity and noise, includes removing all non-green areas, extracting the object with the maximum area, removing empty spaces, resizing it to a rectangle „B" of height 200 pixels and 150 pixels width to maintain the aspect ratio and finally filling all holes. The image is fed into the software which extracts the colour features and Fourier descriptors. Then the image is converted to a gray scale image and then to a binary one with all the leaf pixels to be white (a value of 1) and background pixels to be black (a value of 0) and also extra rows and columns are removed and the images are standardised. Any small voids left after processing due to errors are filled in. Then geometric features dependent on shape of the image are extracted .These are then compared with the ones in database and absolute dissimilarity values as a percentage of the database leaf are calculated for each of the features and summed up. Finally, the absolute variation and percentage variation of each of these parameters from the values in the database is calculated summed. Using the least dissimilarity method, the leaf from the database having the least deviation from the query image is identified as the closest match and result is displayed. The one with least total dissimilarity value is considered to be the closest match to the query leaf and the closest match is displayed as result. Matlab software is used as it has an image processing toolbox which helps in easy implementation of the software. The proposed algorithms is tested with medicinal plant like Thulasi (2 varieties),Thuthuvalai, Karpooravalli, Neem etc.,

VI. RESULT & DISCUSSIONS

First the software enquires about whether the user likes to find a match for the input leaf or he wants to add it to the database. Then it checks the database connection. If the database connection is not enabled, it will pop out an error saying that database connection failed. Else, it will ask the user to input the leaf name. If the file does not exist, then it gives out an error saying that the file does not exist. Else, it reads the image pixels into matrix, and then extracts the colour parameters (new parameter for coloured leaves in code) and Fourier descriptors. It then removes non-green pixels by using a proper threshold. It then standardises the leaf and fills in any holes left in the image. Then it extracts the shape parameters. For a query case, the parameters calculated will be compared with all the leaves in database (using minimum dissimilarity measure) to output the closest matched leaf. For adding the leaf to database, the values will then be sent to a MySQL database where they are stored for future use. The software is trained with 100 (10 number of each plant species) leaves and tested with 50 (tested with different plant species) leaves. During the testing it is observed that 46 leaf images are correctly classified. The efficiency of the implementation is found to be 92%.



View of the display of the result

VII. CONCLUSION

This paper proposes and implements a system for automatic identification of selected medicinal plant leaves from a stored data base of the leaves by processing an image of their leaves. The boundary-based features, moment features & colour of the leaves are used for the purpose of identification of leaf varieties. Such automated classification mechanisms can be useful for efficient classification of plant leaf species. The accuracy of the current method was found to be 92%. Future work would involve, combining venation and texture features to the present shape & colour features to improve recognition accuracy. One major drawback of the implementation is that it is still not very robust to rotation. So, Future work can be directed in that way so as to make the implementation more efficient. The implementation is a little bit computationally expensive. Further optimization of the logic can result in better implementation.

REFERENCES

[1] Zoo Xin, Han Bin, Cheng Jialin, “A measurement approach of leaf area based on digital leaf image processing [J],” Computer Engineering and Applications, vol.27, no.2, pp.194- 196, 2006.

[2] Bai Junhua, Wang Keru, Chu Zhendong, Chen Bing,Li Shaokun, “Study on the measure methods of the leaf area[J], Journal of Shihezi University (Natural Science), vol.23, no.2, April 2005.

[3] Haralick, Robert M., and Linda G. Shapiro, Computer and Robot Vision, Volume I, Addison-Wesley, 1992, pp. 28-48

[4] A. J. M. Timmermans, and A. A. Hulzebosch, “Computer vison system for on-line sorting of pot plants using an artificial neural network classifier”, Computers and Electronics in Agriculture, vol. 15, 1996, pp. 41-55.

[5] X. Gu, J.-X. Du and X.-F. Wang, “Leaf recognition based on the combination of wavelet transform and Gaussia interpolation,” in Proceedings of International Conference on Intel-ligence Computing 2005, ser. LNCS 3644, Springer,2005.

[6] S. Abbasi, F. Mokhtarian, and J. Kittler, “Reliable classification of chrysanthemum leaves through curvature scale space”, Lecture Notes in Computer Science, vol. 1252, 1997, pp. 284-295

[7]. Z. Wang, Z. Chi, D. Feng, and Q. Wang, “Leaf image retrieval with shape feature”, Inter-national Conference on Advances in Visual Information Systems (ACVIS), 2000, pp. 477-487.

[8] Lowe, D.G. (1999) Object Recognition from Local ScaleInvariant Features. ICCV, pp. 1150–1157.

[9]. B.Sathya Bama, S.Mohana Valli, S.Raju, V.Abhai kumar, “Content Based Leaf Image Re-trieval (cblir) using Shape, Color and Texture Features”, in proceedings of Indian Journal of Computer Science and Engineering (IJCSE)

[10] A report [unpublished] on development of software for classification of selected medicinal plant using image processing-Yogesh Kansal, A.Gopal, CEERI Chennai, 2012.

[11] Hu, M.K. 1962. Visual pattern recognition by moment invariants. IRE Trans. Info. Theory. IT-8: 179-187.