

Database Formation for Authentication of Basil (*Ocimum tenuiflorum*) Leaf Using Image Processing Technique

T. Vijayashree ^a and A.Gopal ^b

^aRESEARCH SCHOLAR, Faculty of Electronics, Sathyabama University, Chennai, India

^bCSIR-CEERI, Madras Complex, Chennai, India

e-mail: vijayashreenavajath@gmail.com, agopal@hotmail.com

Abstract

Plants play one of the main roles in our ecosystem. Manual identification for the leaves sometimes leads to greater difference due to look-alike. Hence authentications of leaves are much required for medicinal purposes. The aim of this work is to form the database in order to classify and authenticate tulsi leaf for the purpose of herbal medicines. This paper aims in developing the database with 50 samples of tulsi under various conditions for classification. After forming the database the next step is to compare it with the test leaf to identify the closest match.

Keywords: Authentication, Eccentricity, Principal Component Analysis (PCA), entropy, aspect ratio

1 Introduction

Leaves constitute the most important part of a plant. Leaf identification is the main step for the consideration of the plant. There should be a reliable and efficient leaf recognition system for plant identification. Instead of manual vision, machine vision can be executed for better identification in order to get rid of look-alike leaves. The two main morphotypes cultivated in India and Nepal are green-leaved (Sri or Lakshmi tulasi) and purple-leaved (Krishna tulasi). The image of the leaf is captured using a digital camera with high resolution on a white background under a flat surface. The captured image is fed in to the personal computer for further work. The initial step in image processing is image preprocessing. The purpose of image preprocessing is conversion- converts the color image into gray level image, image enhancement- which enhances the image for better clarity with the region of interest. The second step in preprocessing is morphological operation. It applies a structuring element to an input image creating an output image of the same size. The basic operation of morphological operation is dilation and erosion. Dilation is the process of adding pixels to the boundaries of the object in an image while erosion is the process of removing pixels on boundaries. Adding or removing the pixels depends on the size and shape of the structuring element used to process the image.

The overview of each section is illustrated here. Section 2 deals with the prior and related work, which gives a brief explanation of previous work carried out by various researchers and their results. This section also deals with the techniques used and their advantages and disadvantages. Section 3 deals with the identification method. This section explains the applicable methods for identifying the leaf. Section 4 describes the proposed methodology, and the way it is carried out. This section deals with the optical method, identification technique, parameters used and many more. Section 5 covers the texture analysis, which deals with the texture parameters. Section 6 describes

the database formation procedure. Section 7 discusses the results. Section 8 presents the conclusion and future work.

2 Prior and Related Work

The approach by Lagerwall and Viriri [5] used PCA based texture classification and complex algorithm for the shape feature produced a result of 89.2% in which the author concentrated on the parameters like aspect ratio, compactness, linearity. The morphological feature extraction technique has been used many times. Nikam et al. [7] used a chromatography method which gave an accuracy of 95% who used two techniques in this. One method includes chromatography and the other is based on DNA finger printing. In the former method the leaves were dried under sunlight and shade, then powdered with various solutions like HCl, H₂SO₄, KCl and the powdered sample was taken for testing. Sultana et al. [10] used five features which include shape, texture, color, compactness and aspect ratio of the leaf for analyzing the leaf identification which lead to the problem of adulteration and nomenclatural controversy in misuse of this plant for specific disease.

Kadir et al. [4] approached leaf classification using twelve features like shape, colour, texture, entropy etc., this gave a result of 93.75% for 32 plant leaves which was improved with the original work which gave result of 90.312%. Ehsanirad and Kumar [3] used two methods namely GLCM and PCA which gave an accuracy of 98.46%. Vasantha et al. [12] proposed an approach for leaf identification based on computer aided plant species identification technique. The author used a chain code method and linear computer technique for this method in which the linear technique is used with two sets of dataset for identification. Sambhaji and Andor [9], proposed an approach on leaf recognition using neural network based image processing. The author here used a multilayer perceptron. They used a sample of 9 leafs for recognition the accuracy was about 94%.

Ananthi et al. [1], proposed a technique using pattern recognition of medicinal leaves using image processing techniques. The author used Statistical Package for the Social Science (SPSS) software for determining the dataset. It retrieved the entire value for determining the closest match. The author followed a canny detector for determining the edges and a neural network algorithm. The recognition rate not exceeded between 65-71%. Pallavi and Veena Devi [8], proposed the recognition using Zernike moments. The image of the leaf was resized and the calculation was done. The identification using Zernike moments is somewhat complex for classification. Valliammai and Geethalakshmi [11], proposed a hybrid method for enhancement of plant leaf recognition. The author here used histogram equalization for better enhancement. They have also concentrated on adaptive thresholding where different thresholds are used in different region. They have also taken peak noise to signal ratio (PSNR) for leaf identification. PSNR is defined in logarithmic scale in decibels. It is the ratio of peak signal power to noise power. The evaluation time is also taken for execution. Cope et al. [2], concentrated on the vein characteristics of the leaf using ant colony algorithm. They have done a pixel by pixel operation and also on the level of pheromone, which is the indicator, deposited by ants to signal to other ants the value of the pixel. Lee and Hong [6], implemented a leaf recognition system using leaf vein and shape. They proposed the algorithm using leaf contour extraction. Projection histogram was used to identify the vein and then the features were extracted. The experimental result showed an accuracy of 97.19%.

3 Identification Method

Identifying leaf image was done by various methods:

Chemical Methods - It includes flame tests and chemical tests. Flame and chemical methods are time consuming methods in which the identification of the leaf is done through a long procedure. In the above method it nearly takes fifteen to twenty days for the identification of a single leaf. If the samples are large it nearly takes a minimum of long period for authentication.

Instrumental Methods - It includes spectroscopic technique i.e., mass spectroscopy and thermal leaf which uses a chemo metric technique for identifying the pattern of the leaf to which class it belongs to. Once the method is analyzed a technique using chemo metric is also used for identification.

Optical Methods - It includes digital image processing technique and it is more advantageous than other techniques. In this method the image is converted into gray-scale image and then features are extracted by various techniques to compare the image with the database. In image processing method pattern recognition is considered. The pattern recognition method uses a boundary descriptor which gives the sharp features of the leaf. Vein extraction is taken but tertiary is less reliable.

4 Proposed Methodology

In the proposed methodology identification of leaf is done through image processing algorithm. Tulsi leaves are of various types like karun tulsi, karpooa tulsi and krishna tulsi. Here for the formation of database karun tulsi and karpooa tulsi are used, in which karun tulsi is like purple shaded leaf and Krishna tulsi is green leaf. These are used for various medicinal purpose like ailments, cosmetics, herbal medicines for cold. The leaf is classified and fed in to software which retrieves the related information about the leaf. The proposed network is made to be user-friendly. A step by step process is illustrated in Fig.(1). The image of the leaf is captured with the help of the digital camera with high megapixels of more than about 12 for better clarity. The image is taken under a white background for determining the sharp edges of the leaf. Once the image is captured and fed into the system, image preprocessing is done. Image preprocessing includes the conversion of image from color to gray level. This is done because the color image works with various intensities which will be difficult in programming. In order to overcome this problem the image is converted. The second phase of image preprocessing is image enhancement. In this technique, the image is enhanced for better appearance. The features are extracted which includes low level feature, curvature, image motion, shape, aspect ratio, compactness, entropy, skewness and etc. The third step is pattern recognition which includes pixelization, linear filtering, and quantization. The technique pixelization includes pixel by pixel operation in software. Pattern recognition is done by means of regional descriptors. Edge detection is done using a sobel operator which gives high signal to noise ratio. The last step is software implementation where the captured features are extracted and compared with the data base and the result is obtained. More samples are taken for identifying the parameters to get the exact match with the database. A computer algorithm is used for testing the leaf image, which gives the user friendly relationship between the user and the PC. In this method of analysis various parameters are concentrated for getting a 100% result. The parameters like aspect ratio, entropy, skewness, krutoksis, edge detection, shape, texture, vein features etc. are considered.

The method of venation can also be done for identification of leaves. It uses a set of classifiers for detecting vein. For determining the classifiers a pair of bounds is decided. When the pixel values are within the pair of bounds then it is classified as vein. The leaf with the closest similarity will be considered as the result.

A. Features of Leaf

- i. Shape- The best feature for identifying a leaf. By means of a highly developed algorithm for identification of shape is possible through image processing.
- ii. Venation- It varies in all leaves. It plays the major role for identification of leaves.
- iii. Texture- The other feature for identifying leaf is its texture. It is a powerful regional descriptor that helps in retrieval process. It does not have the capability of finding similar images.
- iv. Color- It is a good identification parameter when there is a variation in color in leaves of different species. But while using color as a feature, it may change due to loss of chlorophyll.

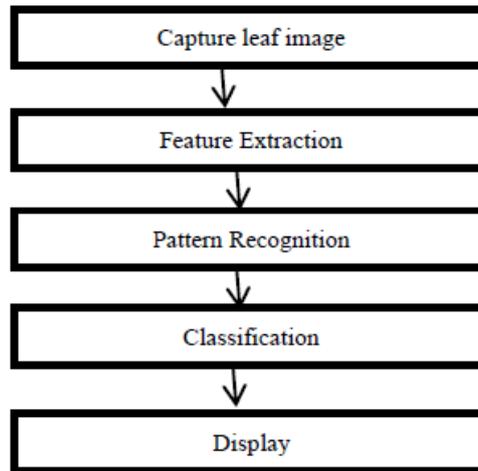


Figure 1. Proposed methodology flow chart.

B. Parameters Used - Geometric Features

AspectRatio - Ratio of width of the leaf to the length of the leaf or it is the ratio of maximum axial length to the minimum axial length of the leaf.

Compactness - Defined as the ratio of the product of area with $4 \times \pi$ to the square of perimeter. It is also referred to as roundness. It is defined as $Compactness = 4 \times \pi \times area / (perimeter)^2$.

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. Dispersion is insensitive to slight discontinuity in the shape such as crack in the leaf.

Centroid - The centroid coordinate of the leaf is obtained and labeled as centroid x and centroid y .

Eccentricity - It is a characteristic feature of any conic section. It is given by $Eccentricity = \sqrt{1 - (b/a)^2}$. Where a and b refer to minimum and maximum axial length respectively for an ellipse.

5 Texture Analysis

Features based on texture are often useful in characterizing the objects and concluding the boundaries between the regions automatically. One measure on texture is based on piece size, which takes the average area of the region in gray level. For determining the texture measure the perimeter and area can be considered. The other method of texture is based on run length. The image is scanned line by line and length in pixels of each threshold. A 2-D texture analysis uses gray-level co-occurrence matrices which show how often each gray level occurs at a pixel located at a fixed geometric location. Texture analysis is categorized into:

Structural - Provides a good symbolic description of image. Powerful tool for structural texture is provided by mathematical morphology. It is based on the arrangement of image primitives such as description of texture based on regularly spaced parallel lines.

Statistical - It is a quantitative measure of arrangement. The approach is based on multidimensional co-occurrence

matrix. The statistical approach is used to characterize the texture of the region as: smooth, coarse and grainy. For considering the statistical moment on texture analysis the moment, mean uniformity, entropy is taken.

Model based - It is not suitable for describing local image structure.

Transform Method - It uses Fourier descriptors, Gabor descriptors and wavelet transform. Gabor filter provides better spatial localization.

Texture content is an important quantity used to describe a region.

6 Database Formation

A class of tulsi leaf is taken here for medicinal quality authentication. Samples of 50 tulsi leaves are taken from various plants. For the formation of the database a set of 50 leaves were taken which includes a sample of 10 karun tulsi leaves. Using MATLAB the database is created, since it has a numerous toolboxes for image processing.

When using image processing toolbox the identification starts with capturing the image. Since here we are dealing with leaf identification morphological processing on image is done which uses image dilation and image erosion. Image dilation uses the structuring element. Dilation adds pixels to the boundaries of objects in an image, while erosion removes pixels on object boundaries. Addition and removal of pixels depend on the structuring element. When image preprocessing is done small hairy like scales present in tulsi leaf will not be seen since the color image is converted into gray-scale image. The geometric features used for the database will be also used for test features which will be done in future. Figure 2, shows the flowchart of the database formation. The initial steps are that the leaves are captured and are fed into the computer. The images are converted into JPEG images before processing. Once the image is transported to JPEG format it is then ready for preprocessing. Figure 3, shows the database formation set. One image from the database is taken and is shown in Fig.(4). The image is converted into gray-level image. The parameters initially taken here is the aspect ratio, compactness, inverse difference moment. The vein parameters will be considered in future for the classification. The leaves vary in different manner with color, size, shape etc. Using software the leaves are coded with a general specification.

The width of the leaf is considered as the reference, the leaves are segmented and the program is coded. Figure 5a, shows the edge detection done for the leaf. Sobel mask is used which has high signal to noise ratio and suppresses noise distortion. Morphological operation is performed such as dilation in order to get the continuity of the edges. The dilated image of the leaf is shown in Fig.(5b). The appearance of the leaf after the removal of holes is shown in Fig.(5c). It is necessary to remove the holes for better result. The segmented portion of the leaf is shown in Fig.(6). The geometric parameters are taken as the column heading for the segmentation process. When chosen sum entropy, inverse moment and other features the result will be better to be compared with the test features. In tulsi leaf the center portion is concentrated since it has the larger area compared with the other parts of the leaf. Once the database is formed it has to be compared with the test samples, by giving the accuracy of the closest match.

7 Result and Discussion

A small part is considered in the database. The above Fig.(5) shows the various stages of segmentation. The database was formed. The future work is test leaves of tulsi to be collected. The leaves are to be compared with the database. The recognized leaf is to be labeled and can be used for medicinal purpose. Histogram processing is done to the leaf which is used for better enhancement. The texture parameters include information about mean difference moment, sum average, sum entropy, sum variance, difference variance. The center portion of the leaf is considered as it has the clear shape and structure. The earlier result of leaf identification with various methods has not achieved a result of 100% accuracy, instead a long time process took place. By this technique 100% accuracy will be achieved as we are using GLCM, PCA, and pattern recognition.

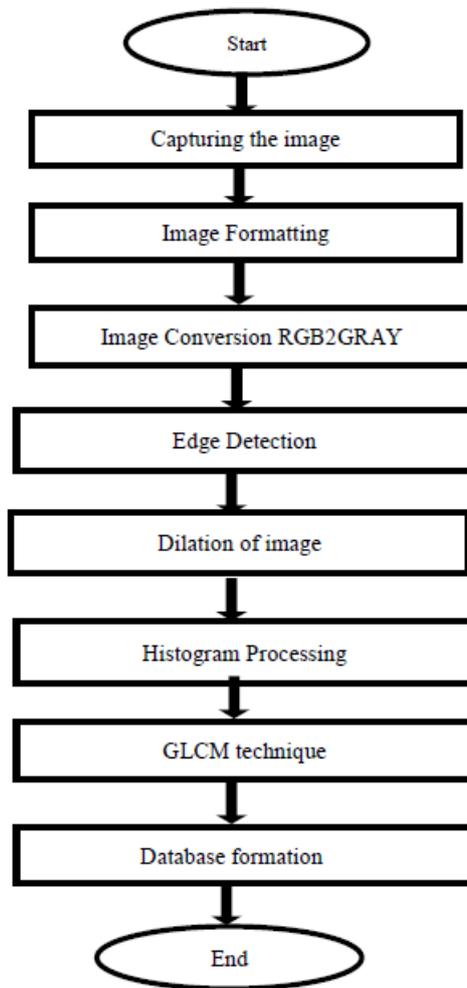


Figure 2. Flow chart for database formation.

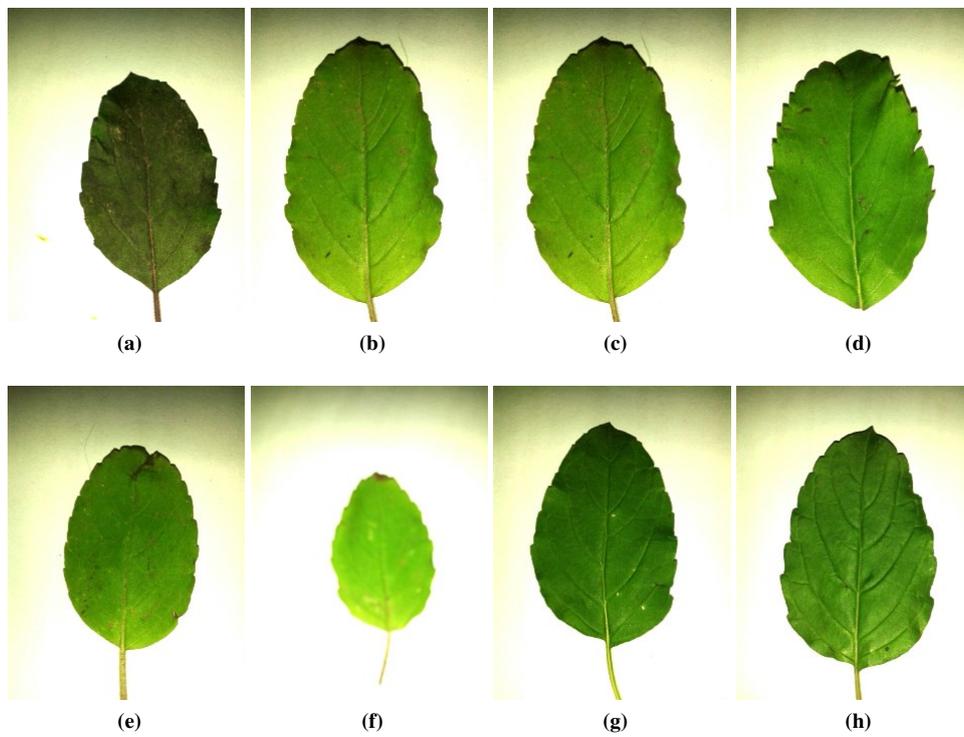


Figure 3. Database.

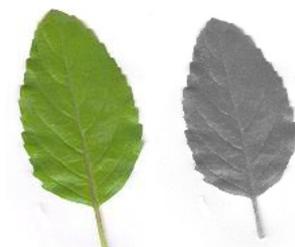


Figure 4. Image preprocessing.

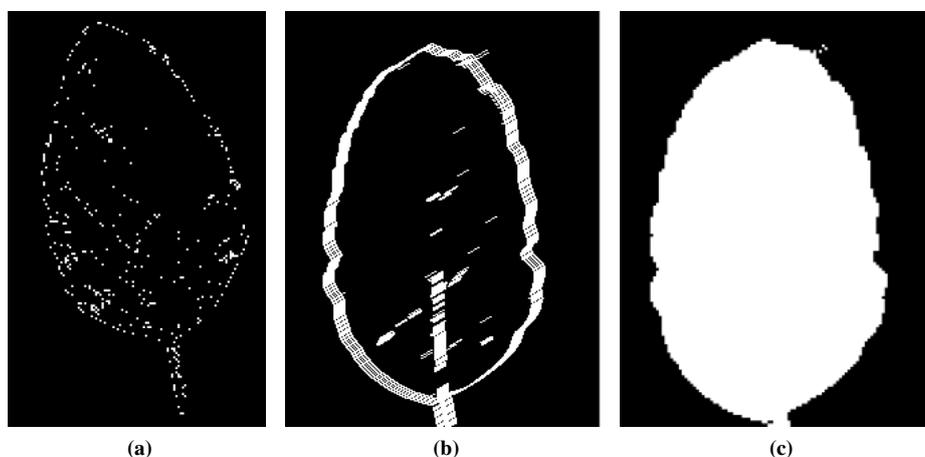


Figure 5. Segmentation procedure (a) Sobel edge detection, (b) dilated image and (c) removal of holes.



Figure 6. Segmented portion.

8 Conclusion and Future Work

The target is to find the most efficient combination of texture features for quick and reliable identification. A larger database will ensure better reliability. For texture analysis shape, texture, feature and venation of leaf are to be tried. The main aim is that it should give a result of 100 % accuracy. The signature parameter of the leaf is the vein. The future work is to study the vein parameter of tulsi leaf. Each leaf has its own signature parameter. For identifying the vein analysis ant colony algorithm can be used. When considering ant-colony algorithm ant like agents is placed at random across the image. These ants move across the image pixel by pixel based on the heuristic evaluation of the pixels. After applying ant-colony algorithm morphological operation can be done. For extracting the leaf vein erosion should be applied after dilation. By using this pattern recognition algorithm it is expected to achieve 100% accuracy.

Acknowledgment

I would like to express sincere thanks to Colonel Dr. JEPPIAAR for his constant encouragement in submitting this paper. A sincere thanks to our beloved Directors Dr. Marie Johnson and Dr. Mariazeena Johnson Sathyabama University for their support in submission of this paper. I wish to thank my supervisor Dr. A.Gopal for his constant encouragement to do this work. I wish to express the sincere thanks to all my family members especially my son Pranav Navajath and my husband T.S.Navajath for their support and motivation in my research work. I thank our Research Head Dr. N.M.Nandhitha for the timely help done by her in developing the database. Firstly and finally I should thank God for being with me throughout.

References

- [1] C. Ananthi, A. Periyasamy, and S. Muruganand. Pattern recognition of medicinal leaves using image processing techniques. *Journal of Nanoscience and Nanotechnology*, 2(1):214–218, 2014.
- [2] J.S. Cope, P. Remagnino, S. Barman, and P. Wilkin. The extraction of venation from leaf images by evolved vein classifiers and ant colony algorithms. In *12th International Conference on Advanced Concepts for Intelligent Vision Systems (ACIVS)*, volume 1, pages 135–144, 2010.
- [3] A. Ehsanirad and Sharath Kumar Y. H. Leaf recognition for plant classification using GLCM and PCA methods. *Oriental Journal of Computer Science & Technology*, 3(1):31–36, 2010.
- [4] A. Kadir, L.E. Nugroho, A. Susanta, and P.I. Santosa. Leaf classification using shape, color and texture features. *International Journal of Computer Trends and Technology*, pages 225–230, 2011.
- [5] R.D. Lagerwall and S. Viriri. Plant classification using leaf recognition. In *22nd Annual Symposium of the Pattern Recognition Association of South Africa*, pages 91–95, 2011.
- [6] K.B. Lee and K.S. Hong. An implementation of leaf recognition system using leaf vein and shape. *International Journal on Bio-Science and Bio-Technology*, 5(2):57–66, 2013.
- [7] P.H. Nikam, J. Kareparamban, A. Jadhav, and V. Kadam. Future trends in standardization of herbal drugs. *Journal of Applied Pharmaceutical Sciences*, 2(6):38–44, 2012.
- [8] V.S. Veena Devi P. Pallavi. Leaf recognition based on feature extraction and Zernike moments. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(2):67–73, 2014.
- [9] E.S. Sambhaji and D.B. Andor. Leaf recognition algorithm using neural network based image processing. *Asian Journal of Engineering and Technology*, 2(2):10–16, 2014.
- [10] S. Sultana, M.A. Khan, M. Ahmad, A. Bano, M. Zafar, and Z.K. Shinwari. Authentication of herbal medicine NEEM (*Azadirachta indica* A.Juss.) by using taxonomic and pharmacognostic techniques. *Pak.J.Bot.*, 43:141–150, 2011.
- [11] N. Valliammai and S.N. Geethalakshmi. A hybrid method for enhancement of plant leaf recognition. *World of Computer Science and Information Technology Journal*, 1(9):370–375, 2011.
- [12] M. Vasantha, V.S. Bharathi, and R. Dhamodharan. Medical image feature, extraction, selection and classification. *International Journal of Engineering Science and Technology*, 2(6):2071–2076, 2010.