Sandalwood Tree Identification using Probabilistic Neural Network as a Classifiers

[1]Nakul Dev, [2] Vivek C,[3] Benaka Santhosh. S
[1]Dept.of ECE,CIT,Ponnampet,Karnataka
[2] Dept.of ECE,CIT,Ponnampet,Karnataka,
[3]Assistant Professor, Dept. of ECE,CIT, Ponnampet, Karnataka,

***Abstract*—Sandalwood (Santalum Album L) is a valuable tree associated with Indian culture. It is the second most expensive wood in the world. Because of which it is at its quick need to be reserved for the future. In order to achieve this, continuous senses of sandalwood trees is required. As an approach, the identification of sandalwood trees using the Probabilistic Neural Network (PNN) as a classifier is implemented here. Several methods to identify plants have been proposed by several researchers. Commonly, the methods did not capture color information because color was not recognized as an important aspect to the identification. The shape, vein, color, and texture features were used here to do the classification of leaf then tree. The result shows that the method of classification gives accuracy of 90%.**

Keywords*—Color features, Sandalwood, PNN, Texture features

I. INTRODUCTION

Santalum album or sandalwood is a small tropical tree and is the most commonly known source of Sandalwood. This species has historically been cultivated, processed and traded since ancient times. Certain cultures place great significance on its fragrant and medicinal qualities. The high value of the species has caused its past exploitation, to the point where the wild population is vulnerable to extinction. So it is at its quick need to be saved. As an approach in this research, the real time sensus of sandalwood trees time to time has been proposed.

Most of the plant identification systems have used the concept of aspect ratio, leaf dent, leaf vein and invariant moment to identify plants. However, all those plant identification systems did not incorporate color information. Color, shape and texture are the common features involved in several applications. But researchers used part of those features only.

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

Neural networks are attracting the researchers in area of pattern recognition because of its power to learn from training data sets. So it is used as classifier on most of plant identification systems. For example, back propagation was used in adaptive route. Selection policy in mobile adhoc networks. PNN is another neural network that has been used in several applications. PNN has proven to be more time efficient than conventional back-propagation based networks and has been recognized as an alternative in real time classification problems.

In this system, we tried to capture the shape, vein, and texture of the leaf of Sandalwood. In implementation, we used Fourier descriptions of PFT, three kinds of geometric features, color moments vein features and texture features based on lacunarity. Then, those features were inputted into the identification system that uses a PNN classifier. Testing was done by using Flavia data set. The result shows that performance of identification system have been improved.

II. FEATURE EXTRACTION

The features of Sandalwood tree leaves are extracted from shape, color, vein and texture. All the features are utilized in the identification system.

A. Shape

If we consider the shape of Santalum album for identification purpose, two features of shape we are going to utilize, they are geometric features and Fourier descriptions of PFT. Geometric features that commonly used in leaf recognition are Slimness and Roundness. Slimness is also called aspect ratio and is defined as,

$$\text{Slimness} = \frac{L1}{L2} \qquad (1)$$

Here l_1 is the width of the leaf and l_2 is the length of the leaf and is as shown in figure 1

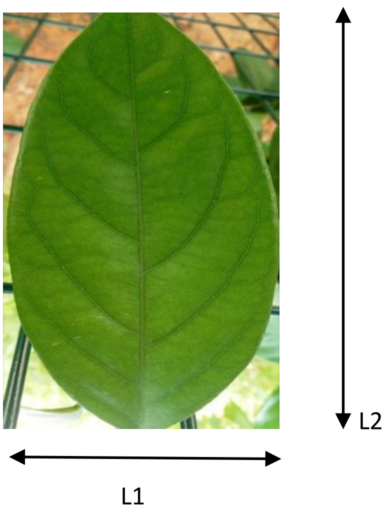


Fig1. Leaf of S.ALBUM

Roundness or compactness is another geometric feature we are going to consider here.

$$\text{Roundness} = \frac{4\pi A}{P^2} \quad (2)$$

Here, A is the area of the leaf image

P is the perimeter of the leaf contour

Polar Fourier Transform (PFT) is very useful to capture shape of a leaf. The descriptions extracted from PFT are invariant under the actions of translation, scaling and rotation as illustrated in figure 2.

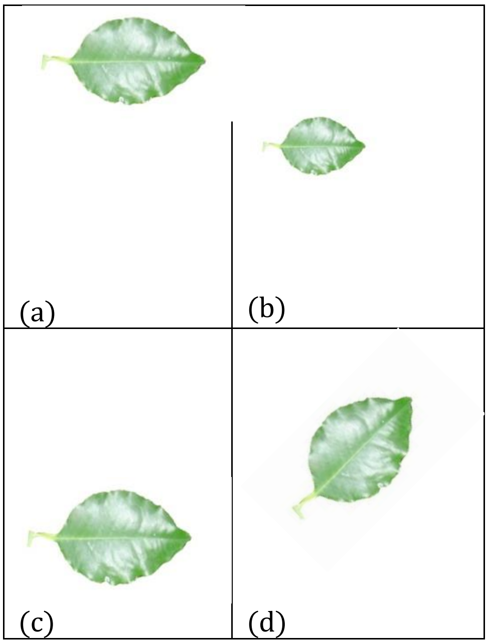


Fig. 2 Translation, scaling, and rotation invariants (a) leaf, (b) change of size,
(c) Change of position, (d) change of orientation

Polar Fourier Transform that is used in this research is defined as

$$PF2(\rho,\phi) = \sum_r \sum_i f(\rho,\phi) exp\left[j2\Pi\left(\frac{r}{R}\rho + \frac{2\Pi}{T}\phi\right)\right] \quad (3)$$

Where,

- 0≤r<R tan θ_i= i(2Π/T) (0≤i<T) ;0≤ ρ<R, 0≤ ϕ<T;
- R is the radial frequency resolution
- T is the angular frequency resolution

Computation of PFT is described as follows,

For example there is an image I = {f(x, y); 0≤x<M, 0≤y<N

In the first step, image is converted from Cartesian space to polar space.$I_p = \{f(r,\theta); 0 \le r < R, 0 < \theta < 2\Pi\}$

Where R is the maximum radius from the centre of the shape. The origin of polar space becomes as centre of space to get translation invariant. The centroid (XC, Yc) calculated by using formula

$$x_c = \frac{1}{M}\sum_{x=0}^{M-1} x\,, \qquad y_c = \frac{1}{N}\sum_{x=0}^{N-1} y \quad (4)$$

In this case (r, ɵ) is calculated by using

$$r = \sqrt{(x-x_c)^2 + (y-y_c)^2}\,, \quad \theta = \arctan\frac{y-y_c}{x-x_c} \quad (5)$$

Rotation invariance is estimated by ignoring the phase information in the co-efficient are retained meanwhile, to get the scale invariance the first magnitude value is normalized by the area of the circle and all the magnitude values are normalized by the magnitude of the first coefficient. So the Fourier descriptions.

$$FDs = \left\{\frac{PF(0,0)}{2\Pi r^2}, \frac{PF(0,1)}{PF(0,0)}, \ldots\ldots, \frac{PF(0,n)}{PF(0,0)}, \ldots, \frac{PF(m,n)}{PF(0,0)}\right\} \quad (6)$$

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

Figure 4 shows the S.Album leave and Table 1 lists the Fourier description of leaf using m=4 & n=6.

Feature	S.Album leaf
1	0.5590
2	0.0024
3	0.1800
4	0.0190
5	0.0770
6	0.4150

Table I

B. Color Features

Color moments of the leaf represent color features to characterize a color image .Features can involved are mean (μ), standard deviation (σ), skewness (ɵ) and kurtosis (γ).

1) Mean— The simple mathematical average of a set of two or more numbers. The mean for a given set of numbers can be computed in more than one way, including the arithmetic mean method, which uses the sum of the numbers in the series, and the geometric mean method

2) Standard Deviation— A quantity expressing by how much the members of a group differ from the mean value for the group.

3) Skewness— It is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point.

4) Kurtosis—It is a measure of whether the data are peaked or flat relative to a normal distribution.

For RGB color space the three features are extracted from each plane R, G and B.The formulas to capture those moments are

$$\mu = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P_{ij} \quad (7)$$

$$\sigma = \sqrt{\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^2} \quad (8)$$

$$\theta = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^3}{MN\sigma^3} \quad (9)$$

$$\gamma = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(P_{ij}-\mu\right)^4}{MN\sigma^4} \quad (10)$$

M and N are the dimensions of image .Pij is value of color on i th coloumn and j th row. The S.Album leaf has the mean, standard deviation, Skewness and kurtosis as given in Table II.

Feature	S.Album
R	μ=72
	σ=22.4763
	θ=5.2656e-005
	γ=-2.9997
G	μ=161
	σ=15.1716
	7.8869e-005
	γ=-2.9997
B	μ=31
	σ=21.8473
	4.2772e-005
	γ=-2.9998

Table II

C Vein Features

Vein features can be extracted by using the morphological features. That operation is performed on the gray scale image with flat disk shaped structuring element. Three features of vein are calculated as follows

$$V1=A1/A, \quad V2=A2/A, \quad V3=A3/A \quad (11)$$

In this case, V1, V2 and V3 represents features of the vein, A1, A2 and A3 represents total pixels of the vein, and A denotes the total pixels on the part of the leaf.

D Texture Features

In this part, we are going to use fractional measure called Lucunarity. The word "lacunarity"refers to a gap or pool as derived from the word for "lake", but in morphological analysis it has been variously defined as gappiness, visual texture, inhomogeneity, translational and rotational invariance etc.

Lucunarity will help to distinguish between two fractals with the same fractal dimension. Definitions of lacunarity are shown as follows

$$L_s = \frac{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N} P_{mn}{}^2}{\left(\frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N} P_{kl}\right)^2} - 1 \quad (12)$$

$$L_a = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|\frac{P_{mn}}{\frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N} P_{kl}} - 1\right| \quad (13)$$

$$L_p = \left(\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\frac{P_{mn}}{\frac{1}{MN}\sum_{k=1}^{M}\sum_{l=1}^{N} P_{kl}} - 1\right)^p\right)^{\frac{1}{p}} \quad (14)$$

Originally these formulae are applied to gray scale images, where Pmn is a gray value at co-ordinate (min). However here only the last formulae or used as texture features. With P having the values 2, 4 and 6. Pmn represents values of color R, G, B of RGB image and intensity in grey scale image. Therefore twelve features are used to represent texture features.

E Data Normalization

Data normalization is a useful step often adopted prior to the use of classifier as a precaution when the feature values vary in different dynamic ranges. In absence of normalization, features with large values have a stronger influence on the cost function in designing the classifier. By normalizing the data, values of all features will be in pre determined ranges. Normalization can be done by using following formulae.

$$Xi = \frac{(X1 - Xmin)}{(Xmax - Xmin)} \quad (15)$$

In this case x^I represents the new value of the features, x1 represents original value of the feature, Xmin is the smallest value of original feature, and Xmax is the largest value of original feature.

III. PROPOSED SYSTEM

The proposed algorithm for S. Album leaf classification followed by tree classification is shown in fig5. In the first step image of leaf is acquired. That leaf image is inputted in to the system for classification. The features contained on the leaf are extracted by Feature Extraction. Then, the features are use by a PNN. The result is an index that represents a plant. Then, plant information getter translates the index into the name of the tree. Before the start of classification, the PNN has been trained once.

Segmentation is used to separate leaf from its background. It can be accomplished by using adaptive threshold .Firstly, an intensity histogram of image is built with 20 bits.Secondly,two major peaks in the histogram that represent the leaf and its background respectively are obtained. Third, find a bit with smallest value that lies between the two major peaks. Then the median of the bits is used as a threshold to separate leaf and its background.

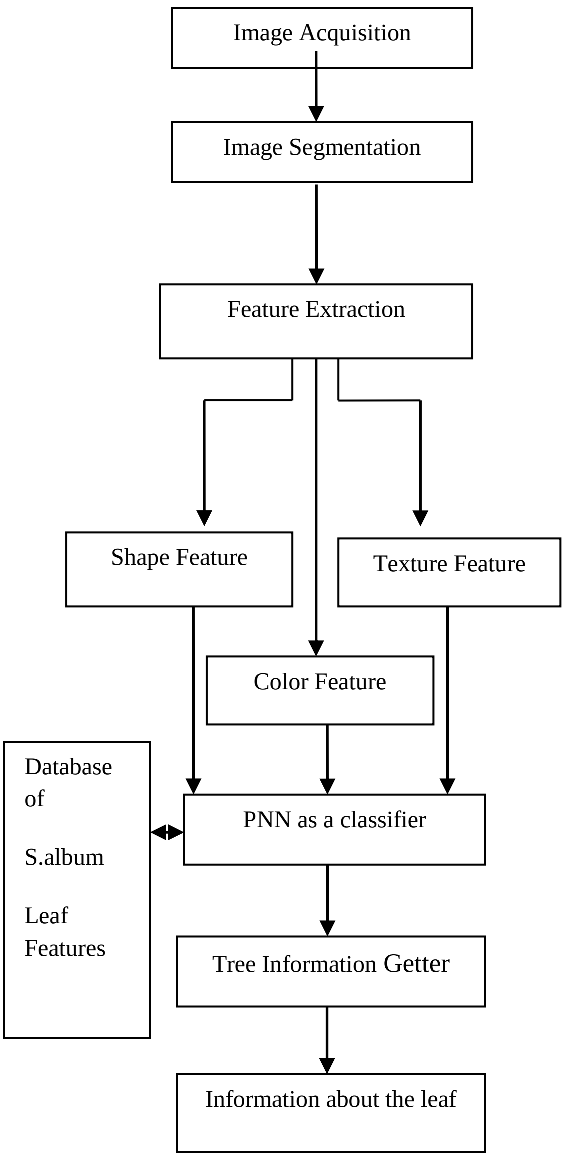
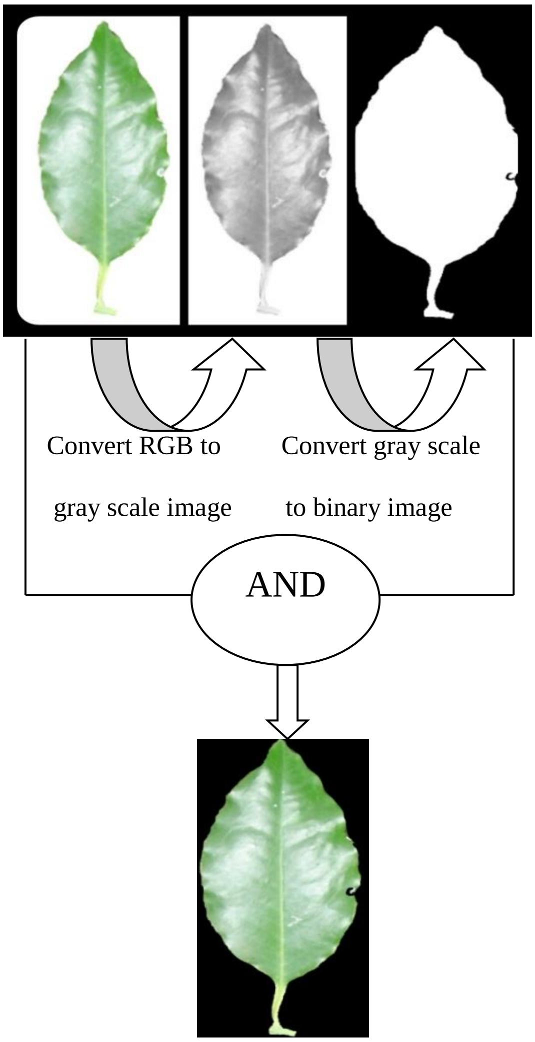
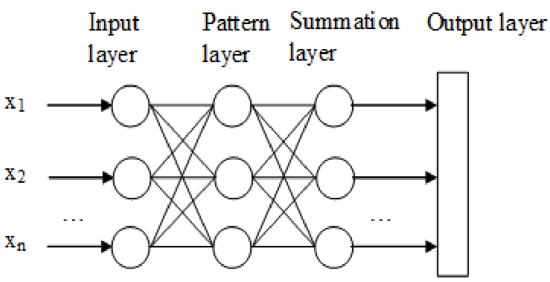


Figure. below shows the process to get area of the leaf. As a first step, the image of the leaf is converted to gray scale image. Then conversion from gray scale to binary form is done by using adaptive threshold. After that several morphological operations are performed to remove holes in the leaf caused by previous thresholding. The leaf is obtained by using operation AND between RGB image and binary image.



The other important part of the identification system is PNN as a classifier.PNN is actually a kind of Radius Basis Function (RBF).Basically PNN classifier adopts Bayer classification rule and density estimation based on Gaussian functions, Figure shows the architecture.



The input layer accepts an input vector. The pattern layer processed the input vector by using weight vector came from training dataset. This layer compute the distances from the input layer to the training input. As a result, a vector that indicate how close the input is to a training input. Then in the summation layer, a vector contains probabilities is found by summing up the contribution for each class. This vector of probabilities is to the output layer. The last layer in PNN structure produces a classification decision in which a class with maximum probabilities will be assigned by 1 and other classes will be assigned by 0.

Mathematically, the probabilities found by summing up the contribution for each class is calculated by using parzen method as follows

$$P(x|w_j) = \frac{1}{(2\Pi)^{d/2}\sigma^d n_j}\sum_{k=1}^{n_j} exp\left(-\frac{(x-X_k)^2}{2\sigma^2}\right) \quad (16)$$

Where $P(x|w_j)$ represents the conditional probabilities 'X' to class w_j, x is input vector, X_k is training data set, 'd' is the number of input vector, n_j is the number of samples for class j, σ is the smoothing factor that its value is input heuristically.

Based on fact that 'x' has class j, if

$P(x|w_i) > P(x|w_j)$, $i \neq j$

then $P(x|w_j)$ can be calculated as follows,

$$P(x|w_j) = \frac{1}{n_j}\sum_{k=1}^{n_j} \exp\left[-\frac{(x-X_k)^2}{2\sigma^2}\right] \quad (17)$$

IV. RESULTS

The image of the purticular region having the sandalwood trees have been captured using high definition camera. From the acquired image, each individual tree, followed by some 10 samples of leaves in that tree are extracted using feature extraction of image segmentation. Each leaf is compared in all the characteristics provided in last section. With the database of S.Album leaf provided. Depending on this the tree type is detected. In this case,PNN classifier is adjusted by using smoothing factor as equal 0.05

In order to get the perfomance of the system, the following formula is used

$$\text{Perfomance} = \frac{n_r}{n_t} \quad (18)$$

where n_r is relevant number of images and n_t is the total number of query.

Table below shows the results. As given in the table, combination of shape, color (without kurtoris), vein, and texture features gives the best result with accuracy of 90%. Based on the results, we can see that all kinds of features have important contributions except the kurtoris.

Features	Perfomance
PFT	74.6875%
PFT + 3 geometric features	77.5000%
PFT + 3 geometric features + mean of colors	82.5000%
PFT + 3 geometric features + mean of colors + standard deviation of colors	88.1250%
PFT + 3 geometric features + mean of colors + standard deviation of colors + skewness of colors	88.7500%
PFT + 3 geometric features + mean of colors + standard deviation of colors + skewness of colors +kurtosis of color	87.8125%

V. CONCLUSIONS

An algorithm for sandalwood tree identification has been developed. This method in corporates shape, vein, color and texture features, and uses PNN as a classifier. Fourier descriptors, slimness ratio, roundness ratio and dispersion are used to represent shape features. Color moments that consists of mean, standard deviation, and skewness are used to represent color. Twelve textures. The result gives up to 90% of accuracy. In the future work, we are planning to implement the system using some more advanced classifiers.

REFERENCES

[1] 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, 2006.

[2] S. G. Wu, F. S. Bao, E. Y Xu, Y-X. Wang, Y-F. Chang, & Q-L. Xiang, "A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network", *IEEE 7th Interantional Symposium on Signal Processing and Information Technology*, Cairo, 2007.

[3] K. Singh, I. Gupta, & S. Gupta, "SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 3 (4), pp. 67-78, 2010.

[4] J-X. Du, D-S. Huang, X-F. Wang, & X. Gu, "Computer-aided Plant Species Identification (CAPSI) Based on Leaf Shape Matching Technique", *Transactions of the Institute of Measurement and Control* , vol. 28 (3), pp. 275-284, 2006.

[5] P. Hiremath, & J. Pujari, "Content based Image Retrieval based on Color, Texture and Shape Features Using Image and Its Complement", *International Journal of Computer Science and Security* , vol. 1 (4), pp. 44-50, 2011.

[6] B. Jyothi, Y. M. Latha, & V. Reddy, "Medical Image Retrieval using Multiple Features", *Advances in Computational Sciences and Technology* , vol. 3 (3), pp. 387-396, 2010.
[7] S. Theodoridis, & K. Koutroumbas, "An Introduction Pattern Recognition", Burlington: Academic Press, 2009.
[8] A. Kulkarni, "*Artificial Neural Networks for Image Understanding",* New York: Van Nostrand Reinhold, 1994.
[9] M. Mercimek, K. Gulez, & T. V. Mumcu, "Real Object Recognition Using Moment Invariants", *Sadhana* , vol. 30 (6), pp. 765-775, 2005
[10] Z. Zulkifli, "Plant Leaf Identification Using Moment Invariants & General Regression Neural Network", Master Thesis, Universiti Teknologi Malaysia, 2009.
[11] A. Kadir, L.E. Nugroho, A. Susanto, & P.I. Santosa, "A Comparative Experiment of Several Shape Methods in Recognizing Plants", *International Journal of Computer Science & Information Technology (IJSIT)* , vol. 3 (3), pp. 256-263.
[12] D. Zhang, "Image Retrieval Based on Shape", Unpublished Dissertation, Monash University, 2002.
[13] I. Kunttu, L. Lepisto, & A. Visa, "Image Correlogram in Image Database Indexing and Retrieval", *Proceedings of 4th European Workshop on Image Analysis for Multimedia Interaction Services*, pp. 88-91, London, 2003.
[14] Q-K. Man, C-H. Zheng, X-F. Wang, & F-Y.Lin,"Recognition of Plant Leaves Using Support Vector", *International Conference on Intelligent Computing*, pp. 192-199, Shanghai, 2008.
[15] 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.
[16] T. Acharya, & A.K. Ray, "Image Processingg Principles and Applications", New Jersey: John Wiley & Sons, Inc, 2005.
[17] Y.A.O. Min, Y.I. Win-Sheng,S. Bin, &D.A.I. Hong-Hua, "An Image Retrieval System Based on Fractal Dimension", Journal Zheijang University Science, vol. 4(4), p. 421-425, 2003.