Plants Identification by Leaf Shape using GLCM, Gabor Wavelets and PCA

Akshay A Patil[#1], K. S.Bhagat[*2]

[1]M.E.(E&TC) Student, J .T. Mahajan college of Engineering, Faizpur. INDIA.

[2] M.E.(E&TC) Faculty, J .T. Mahajan college of Engineering, Faizpur. INDIA.

***Abstract**: Plants are essentials for life on Earth. Different species of plants can be distinguished with the help of leaf shapes, petals barks, and fruits. A digital recognition of plant species is a now a days in demand for various purposes. A new method for recognizing and identifying plants has been devised. Leaf images are pre-processed to remove noise using median filtering on different image planes separately. Gabor wavelet and Gray-Level Co-Occurrence Matrix (GLCM) texture features are used to recognize/identify leaf shape, for classification Decision Tree Classifier is used. We found improvement in results by utilizing combination of GLCM and Gabor features. Effective discriminative power is increased by dimensionality reduction with the help of Principal component analysis (PCA) and Decision tree as a classifier. This intelligent system provides accurate results within in less time by utilizing photographs of plants leaves, to identify tree species. The real time database is prepared for the experimental use. The database contains various leaves with various shapes, colours and size. Experiment is carried out with the different leaves of different classes and tested results are quite good except real time database.*

Keywords— *Plants, leaf shape, digital image processing, GLCM, Gabor wavelets, Principal component analysis, decision tree.*

I. INTRODUCTION

India is an agriculture based country, where 70% of the population depends on agriculture [1]. Now-a-days the people of India wants to use some technology that make them work easier, faster and with more perfection and with less cost. Plant classification remains very useful and important task for scientist, field guides and others. Similarly plants physical features like texture weight, color and hardness should be taken into account for better recognition.[2].

II. RELATED WORK

Modern plant taxonomy starts with the Linnaeus' system of classification [3]. This is a plant classification and nomenclature system and is currently used, albeit in a revised version. Gu et al [4] used a combination of wavelet transform and Gaussian interpolation to extract the leaf venation and contour of leaf from its image and derived run length features based on the image. The plants were then classified based on the extracted and derived features using 1-NN, k-NN and radial basis probabilistic neural network (RBPNN). Zernike Moment Invariant (ZMI) are better when combined with other classifiers as shown by Abdul[5].ZMI combined with Gray Level Co- occurrence Matrices, geometric features, color moments accuracy obtained is 94.96%. In the paper [6] GLCM and PCA techniques were utilized on 390 leaves to classify 13 types of plants. Results for GLCM were 78% accuracy and for PCA was 98%. Computation time for GLCM is less than PCA. T.Vijayshree and A. Gopal [7] experimented on dataset with 50 samples of tulsi under various conditions for classification. They developed a system consist of GLCM, PCA and pattern recognition, achieved 100% accuracy. Zheru chi et al [8] proposed a novel method of Gabor filter bank classifiers. In the pre-processing phase Gabor filters are employed by Yusof et al[9] with GLCM to extract features, a multilayer neural network is used for classification. Sablatnig [10] proposed GLCM and wavelet features for identification of tree species from bark images, leave images and needles, using SVM as classifier. Classification rate of bark images was 69.7%. G.Beliakov et al [11] discussed the problem of texture recognition based on the gray level co-occurrence matrix(GLCM). Z. Le-Qing et al [12] proposed a novel insect recognition algorithm based on color histogram and GLCM which can identify insect species from low resolution wing images. The recognition rate is as high as 71.1%.An ideal time performance is also achieved. The experimental result testifies the efficiency of proposed method. M. Mustafa et al [13] proposed a time-frequency approach or spectrogram image processing technique for analyzing EEG signals. Gray Level Co-occurrence Matrix (GLCM) texture feature were extracted from spectrogram image and then Principal component analysis (PCA) was employed to reduce the feature dimension. The purpose of them was to classify EEG spectrogram image using k-nearest neighbour algorithm (KNN) classifier. The result shows classification rate was 70.83% for EEG spectrogram image.

III. MATERIALS AND METHODS

In computer image analysis one of the important topic is study of texture recognition and classification. Computer image analysis has extensive range of applications in several fields, such as food processing, medical science and remote sensing. A variety of methods for texture classification have been projected in the literature [14]. one of the most dominant method for general texture classification known as the Grey Level Co-occurrence Matrix (GLCM) [15] is employed for texture features extraction techniques.

1. Grey Level Co-Occurrence Matrix (GLCM): It's texture size have been the workhorse of image texture since they were proposed by Haralick in the 1970s. As broad-spectrum, GLCM could be computed as follows. First, an original texture image D is re-quantized into an image G with reduced number of grey level, Ng. A typical value of Ng is 16 or 32. Then, GLCM is computed from G by scanning the intensity of each pixel and its neighbor, defined by dislodgment d and angle ø. A dislodgment, d could take a value of 1,2,3,…n whereas an angle, ø is limited 0^0, 45^0, 90^0 and 135^0 . Decision tree classifier takes input as features from GLCM computation, or from Gabor wavelets.

$$\text{Energy}: \sum_{i,j} P(i,j)^2 \qquad (1)$$

$$\text{Entropy}: -\sum_{i,j} P(i,j)\log P(i,j) \qquad (2)$$

$$\text{Homogeneity}: \sum_{i,j} \frac{1}{1+(i-j)^2} P(i,j) \qquad (3)$$

$$\text{Inertia}: \sum_{i,j} (i-j)^2 P(i,j) \qquad (4)$$

$$\text{Correlation}: -\sum_{i,j} \frac{(i-\mu)(j-\mu)}{\sigma^2} P(i,j) \qquad (5)$$

$$\text{Shade}: \sum_{i,j} (i+j-2\mu)^3 P(i,j) \qquad (6)$$

2. Gabor wavelets: Dennis Gabor invented Gabor wavelets by with complex functions constructed to serve as a base for Fourier transforms(harmonic function) in information theory applications. They are extremely related to Morlet wavelets. 2-D Gabor filters can be formed into 2-D Gabor wavelets as Gaussian function varies in dilation and the harmonic function varies in rotation and frequency. Put another way, the uncertainty in information carried by this wavelet is minimized. Because of Non-Orthogonality (the down side), it is difficult to obtain sufficient decomposition at basis level. Since their initiation, different applications have appeared, from image processing to analysing neurons in the human visual system.

Remaining of this paper is organised as follows. In section II related work of some authors is described. section III describes basic theories of GLCM , Gabor wavelets, PCA and decision classifier. In section IV experimental results are presented. Section V concludes with future scopes.

A FEATURE EXTRACTION

In general pattern recognition systems, there are two steps in building a classifier: training and testing (or recognition). The pattern recognition involves following three steps (1) Pre-processing (2) Feature Extraction (3) Classification. These steps can be further broken down into sub-steps.

Training:

1. Pre-processing: Process the data so it is in a suitable form.
   Here we have used median filter to remove noise from the
   leaf images.
2. Feature extraction: Extract the features from the training leaf images such as GLCM and wavelet features.
3. Dimensionality Reduction: By applying Principal Component
   Analysis (PCA) on the extracted features. Reduces the
   amount of data by extracting pertinent information.
4. Model Estimation: From the finite set of feature vectors, need to estimate a model (usually statistical) for each class of the training data. Here we have constructed the decision tree model for leaf recognition.

Recognition:

1. Pre-processing: Process the test data so it is in a suitable form. Here we have used median filter to remove noise from the leaf images.
2. Feature extraction: Extract the features from the test leaf images such as GLCM and Gabor wavelet features.
3. Dimensionality Reduction: Apply Principal Component Analysis (PCA) on the extracted features. Reduce the amount of data by extracting relevant information.
4. Classification: Compare feature vectors to the various models and find the closest match. One can match the feature vectors obtained in training set. Here we have used the constructed decision tree model for prediction of the leaf
   class.

B Classification (Recognition)

Once the features have been extracted, then these features are to be used to classify and identify the leaf using decision tree classifier to classify plants based on texture-related features of leaf such as GLCM features like contrast, correlation, energy and homogeneity and wavelet features like Gabor.

1. Classification Based on Decision Trees

Decision tree build categorization or deterioration models in the form of a tree construction, by tree generation algorithms. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data. To the downside, 1) Overlap due to which search time and memory space increases. 2)There is accumulation of errors from level to level in large trees. 3) Well designed tree gives better performance than poorly designed tree.

IV. EXPERIMENTATION

Three different datasets have been collected for comparison of proposed method. All the experiments are programmed by MATLAB and run on Intel(R) core i-3 with the clock of 2.40 GHz and the RAM of4GB under windows 7 environment. Experimental Result for UCI machine Learning Repository Leaf Dataset.

A. Results for UCI Machine Repository Dataset

Table 1

Details about leaf samples of different types of plants and the result of leaf recognition for UCI Machine Repository Dataset

Class	Common Name	No. of test leaf samples	No. of correctly recognized images Using GLCM	Using Gabor wavelet	Combined Features
1	Quercussuber	12	12	12	12
2	Salix atrocinera	13	13	13	13
3	Populusnigra	16	16	14	16
4	Alnus sp.	12	10	11	12
5	Quercusrobur	13	8	10	10
6	Cratae-gusmonogyna	12	8	12	11
7	Ilex-aquifolium	10	10	10	8
8	Nerium-oleander	10	7	10	8
9	Betula-pubescens	5	5	5	5
10	Tiliato-mentosa	12	10	10	11

Table 2

Classification Accuracy for UCI Machine Repository Dataset

Sr No.	Features	Percentage of Correct Classification	Percentage of Incorrect Classification
1	GLCM	77.652370%	22.347630%
2	Gabor Wavelet	85.101580%	14.898420%
3	Combined Features	88.036117%	11.963883%

B. Results for Swedish Leaf Dataset

Table 3

Details about leaf samples of different types of plants and the result of leaf recognition

Class	Common Name	No. of test leaf samples	No. of correctly recognized images Using GLCM	No. of correctly recognized images Using Gabor wavelet	Using Combined Features
1	Ulmu-scarpinifolia	75	75	75	75
2	Acer	75	75	74	75
3	Salix aurita	75	65	69	71
4	Quercus	75	71	74	73
5	Alnusincana	75	72	75	71
6	Betula-pubescens	75	74	69	73
7	Salix alba 'Sericea'	75	72	71	73
8	Populu-stremula	71	68	65	69
9	Ulmusglabra	75	72	70	73
10	Sorbus-aucuparia	75	67	69	70

Table 4
Classification Accuracy for Swedish Leaf Dataset Dataset

Sr No.	Features	Percentage of Correct Classification	Percentage of Incorrect Classification
1	GLCM	94.647636%	5.352364%
2	Gabor Wavelet	94.112400%	5.887600%
3	Combined	96.074933%	3.925067%

C. Experimental Results for Real Time Dataset

We have evaluated the performance of the proposed system for real time image dataset. We got the accuracy of 32.25%. The Fig 1 shows the result for the real time image recognized.

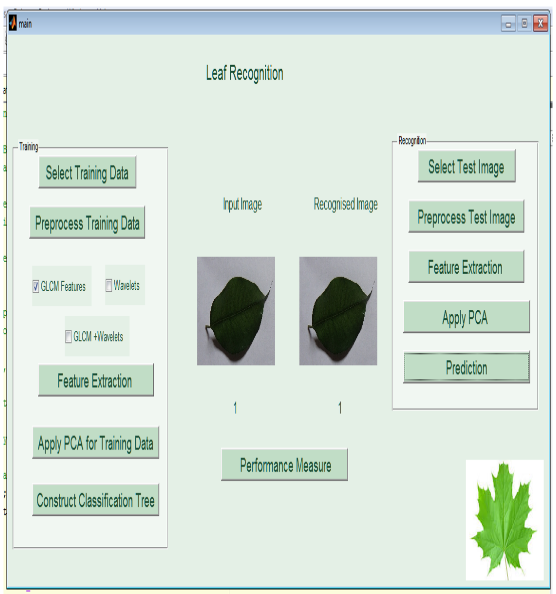

Fig 1. Real time image recognized

V. Conclusion and Future Scope

In this paper, the classification based on the recognizing the leaves images with extracted texture features was proposed and performed. Various features related to the texture of the leaves were studied and the most appropriate features were used for leaf image-based plant classification. The features selected in this automated plant identification were: 1) Gabor wavelet Features, 2) GLCM Features and 3) Combined Gabor wavelet and GLCM features. The texture features have been extracted with using the combined feature matrix of Gabor wavelet features and Gray-Level Co-occurrence Matrix (GLCM) and the Principal Component Analysis (PCA) algorithms, on various datasets. The experimental result shows that the proposed method of combined features performs better for classification of leaf images. The proposed method could used to integrate both simple and complex leaves for plant identification. Developed system could be used to identify medicinal plants for particular diseases of human beings.

REFERENCES

[1] www.agriinfo.in

[2]Akshay A Patil, K. S.Bhagat,"Plants Identification by Leaf Shape using GLCM ,Gabor Wavelets and PCA.," (communicated and accepted) International Journal of Engineering Trends and Technology (IJETT)

[3] M. J. Dallwitz, "A general system for coding taxonomic description". [online]. vol. 29. Available:http://delta-intkey.com

[4] X. Gu et al.,. Wang, "Leaf recognition based on the combination of wavelet transform and Gaussian interpolation," ICIS, vol. 3644/2005, pp. 253-262, 2005.

[5] Abdul,K., Lukito,E.N., Adhi,S., and Santosa,P.I,. "Experiments of zernike moments for leaf identification". *Journal of Theoretical and Applied Information Technology*, 41(1),pp.83-93, 2012.

[6] A. Ehsanirad and Sharath kumar Y. H., "Leaf recognition for plant classification using GLCM and PCA methods", Oriental Journal of computer Science & Technology vol. 3(1), pp. 31-36, 2010.

[7] T.Vijayshree and A Gopal, "Database Formation for Authentication of Basil Leaf using Image processing Technique,"Gate to Computer Vision and Pattern Recognition,vol.1,no.1,pp.9-17,2015.

[8] Zheru Chi, Li Houqiang and Wang Chao, "Plant species recognition based on bark patterns using novel Gabor filter banks",Proceedings of International Conference on *Neural Networks and Signal Processing*, vol. 2, pp. 1035 – 1038, 2003.

[9] R. Yusof, N.R. Rosli and M. Khalid, "Using Gabor Filters as Image Multiplier for Tropical Wood Species Recognition System", 12th International Conference on *Computer Modelling and Simulation* (UKSim), pp. 289–294, 2010.

[10] Stefan Fiel and Robert Sablatnig, "Automated identification of tree species from images of the bark, leaves and needles," 16th Computer Vision Winter Workshop Andreas Wendel, Sabine Sternig, Martin Godec (eds.) Mitterberg, Austria, February 2-4, 2011.

[11] G.Beliakov, S.James,L. Trioia,"Texture recognition by using GLCM and Various Agregation Function," IEEE FUZZ,pp.1472-1476, 2008.

[12] Z.Le-Qing,Z. Zhen,"Auto- classification of Insect Images Based on Color Histogram and GLCM ," IEEE FSKD, vol.6,pp.2589-2593, 2010.

[13] M.Mustafa, M.Taib,Z.Murat, N.Hamid," GLCM Texture classification for EEG Spectrogram Images," IEEE IECBES, pp.373-376, 2010.

[14] R.M. Haralick, Statistical and Structural Approaches to Texture, Proc. IEEE, 67, 1979, pp.786-804.

[15] C.C. Gotlieb, and KH.E. Kreyszig, Texture Descriptors based on Co-occurrence Matrices, Computer Vision, Graphics, and Image Processing, 51, 1990, 70-86.