An Efficient Segmentation and Classification Technique for Automatic Plant Identification

N.Valliammal
Assistant Professor, Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore-641043
Tamilnadu, India.

Dr.S.N.Geethalakshmi
Professor, Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore-641043
Tamilnadu, India.

Abstract— **The field of leaf recognition for plant classification has experienced an increased need for fast and efficient classification algorithms to aid in keeping track of important plants. Plant identification through leaf recognition is a research field that has captured the attention of many botanist for several years. This paper describes the plant recognition through segmentation and classification of leaf images, it is essential for quantification of outlined leaf structures and for visualization of relevant image data. Segmentation in this area includes plant identification and plant classification. The ultimate goal of our approach is to develop a system where user can easily process a unknown plant leaf with efficient image processing and computing technique within fraction of seconds. The effectiveness of the system is proved by quantitative approach. The performance of the proposed work is compared with the existing traditional classification algorithm and real time unknown plant leaf image.**

Keywords-Segmentation;Classification; Plant Leaf Identification; K Means Clustering; Wavelet;Feature Extraction;BPNN.)

I. INTRODUCTION

Globally, it has been found that there are more than 1.7 million species of living organisms (human beings, plants and algae) on Earth, out of which, plants species plays a vital role in human life. Plants are an essential resource for human well-being and can exist everywhere. Most of the plants carry significant information for the development of human society and are considered as essential resource for human well-being. Plants are of plenty of use as they form the base for food chain and a lot of medicines are derived from plants. Plants are also vitally important for environmental protection.

Two main plant aspects of plant taxonomy that play a vital role in these endeavours are the identification and classification of plants.

- *Plant Identification* is the determination of the identity of an unknown plant in comparison with previously collected specimen. The process of recognition connects the specimen with a botanical name. Once this connection is established, related details like name and other properties of the plant can be easily obtained.
- *Plant Classification* is the placing of known plants into groups or categories to show some relationship. They use features that can be used to group plants into a known hierarchy.

This research focuses on the automation of plant identification through leaf recognition. Apart from using the whole plant, the automation of plant identification can be performed using various parts of a plant anatomy like stem, flower, petal, seed and leaf.

This study uses the leaf part of the plant to identify a plant. The continued interest in biodiversity along with the ease of creating digital images, increased the need for processing power of computers and economical methods. In order to gather the information, plant identification using computers has become an interesting subject of research. Global shortage of expert taxonomists has further increased the demand for automated tools that would allow non-botanical persons to carry out valuable field work of identifying and characterizing plants. These tools are of importance in several fields including agriculture, forestry and pharmacological science (Cotton Incorporated USA, 2009; National Institute for Agricultural Botany, 2005). The first step during the design and development of such tools starts with leaf recognition. Compared with other methods, such as cell and molecule biology methods, identification of plants based on leaf image is the most successful and proven method (Wu *et al.*, 2007). Sampling leaves and obtaining a photograph of them is convenient and viable, due to the availability of low cost digital cameras.

Currently, plant identification through leaf recognition involves finding information about a plant that most matches the species name (key) that has to be known in advance. Though identifying plants using such key is a time consuming task, correct utilization of key plays a direct role in the success of the plant search. The alternative method of allowing users to provide a leaf image is very convenient, user friendly and eliminates the need for key. The task combines the challenges of different fields like image processing, machine learning and pattern recognition. Identifying the most favourable algorithms and techniques from these fields, for plant identification through leaf recognition, is the main focus of this study.

The proposed method is organized as follows: Section I provided a brief introduction to the topic. Section II discusses the need for image segmentation. Section III presents the Image feature extraction and classification algorithm for leaf Image recognition. The simulation results with different parameter evaluation are presented in section IV. Finally, conclusions are given in section V.

II. SEGMENTATION

In general imaging science, segmentation is an important step in any image analysis process where an image is taken as input and the output is some detailed description of the scene or object. It has the same importance in leaf recognition also. Figure 5.1 depicts the steps involved in typical image analysis workflow showing segmentation as key step for succeeding image representation and recognition stages.

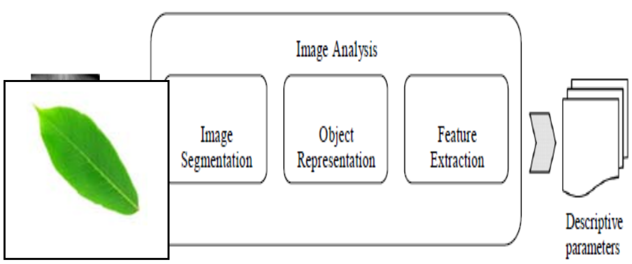

Figure 1 : Steps in Image Analysis

Classically, image segmentation is defined as the partitioning of an image into non-overlapping, constituent regions which are homogeneous with respect to some characteristic such as intensity or texture.

A.Proposed segmentation method

The proposed algorithm is used to segment the leaf image from its background uses texture and color features to form a feature vector, which are then segmented using K means clustering algorithms. The algorithm makes use of wavelet frame decomposition during the extraction of texture features and L*u*v* color space to extract color features. The Wavelet based Segmentation using Clustering and Texture based Color Features WCF method consists of five major steps and the algorithmic flow is presented in Figure.

1. Extraction of *texture features*
2. *Extraction of color features*
3. *Creation of high dimensional feature space*
4. *Perform mean shift filtering to smooth and preserve edges*
5. *Use of K-Means clustering* to obtain segments

The algorithm begins by converting the input leaf image from RGB color space to L*u*v color space. This process produces the results with three color channels, L, u and v respectively. Simultaneously, the algorithm also applies wavelet transformation to obtain the four subbands LH (Low High), HL(High Low), LH(Low High) and HH(High Low). The next step classifies the coordinates into four texture classes for which the median energy is calculated. The combined texture and color features form a high dimensional feature space, which is converted to a low level space using a mean shift filtering algorithm. This result is then grouped by K-Means algorithm to segment the input image into ROI and background. The ROI is the output of the extracted leaf image.

Thus, the sequential steps of WCF method for segmenting the leaf image from its background can be summarized as follows.

Step 1 : Convert image from RGB color space to L*u*v color space and decompose each color channel L, u and v.

Step 2 : Obtain gray scale image from RGB color space.

Step 3 : Use wavelet frame transformation to decompose the gray scale image into sub-bands (LL, LH, HL, and HH) and use the LH and HL subbands for further processing.

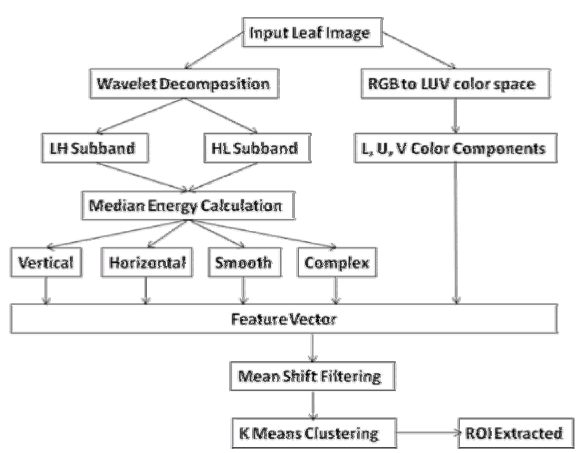

Figure 2 : Proposed WCF Method

Step 4 : Texture Feature Extraction

(a) Calculate the median energy for LH and HL subbands coefficients in a local window. The size of the window should be large enough to capture the local texture characteristics.

(b) Use K-means clustering algorithm to classify the energy values in two classes for each subband.

(c) Using these information, classify the pixels in textured region into four groups, namely, smooth, vertical, horizontal and complex.

Step 5 : Generate the feature vector such that every pixel in the image has p-dimensional feature vector which includes spatial (x,y), color (gray-level or L*u*v values) and texture (smooth, vertical, horizontal or complex) information.

Step 6 : Filter the image using mean shift algorithm in higher dimensional feature space, which includes spatial, color and texture information. The filtering operation can be controlled by setting the spatial window radius (hs) and color range (hr). The filter output (convergence point in mean shift algorithm) is determined by color as well as texture information unlike in standard mean shift filtering. This provides better discrimination between regions where colors are similar but texture is different.

Step 7 : Segment the output image using K-Means clustering algorithm.

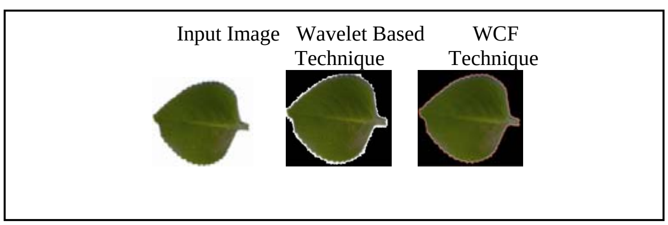

Figure 3 : Proposed segmented Method

III. FEATURE EXTRACTION NAD CLASSIFICATION

After enhancing and extracting the leaf image, the next step of CAP-LR is feature extraction and selection. The main aim of this step is two folds.

Feature Extraction : Converts the image data into a representation that allows comparisons between leaf images by extracting leaf properties .

The task of the feature extraction and selection methods is to obtain the most relevant information from the original data and represent that information in a higher dimensionality space.

Texture of a plant may be due to having many veins in different directions or parallel lines of different colors. Classical Gabor filters gives rise to important difficulties when implemented in multiresolution space (Zhang *et al.*, 2004; Fischer *et al.*, 2006). The texture features extracted from the segmented leaf images are energy, entropy, homogeneity and variance.

The energy of an image is calculated as described below. To calculate energy (also called Uniformity) first the Angular Second Moment (ASM) is to be calculated. Both ASM and Energy use each P_{ij} as a weight for itself. High values of ASM or Energy occur when the window is very orderly.

$$\text{ASM equation} = \sum_{i,j=0}^{N-1} P_{i,j}^{2}$$

Energy is now calculated as the square root of the ASM as shown in Equation .

$$\text{Energy} = \sqrt{\text{ASM}}$$

The entropy is calculated using the formula

$$\text{Entropy} = \sum_{i,j=0}^{N-1} P_{i,j}(-\ln P_{i,j})$$

The Homogenity feature is calculated as

$$\text{Homogeneity} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$$

Homogeneity is the most commonly used measure that increases with lesser contrast in the image window.

The variance for the horizontal and vertical directions is calculated as below.

$$\text{Variance} = \sigma_i^2 = \sum_{i,j=0}^{N-1} (i-\mu_i)(P_{i,j}) \quad ,$$

$$\sigma_j^2 = \sum_{i,j=0}^{N-1} (j-\mu_j)(P_{i,j})$$

Image classification is an area in image processing where the primary goal is to separate a set of images according to their visual content into one of a number of predefined categories. It is the problem of finding a mapping from images to a set of classes, not necessarily object categories. Each class is represented by a set of features (feature vector) and the algorithm that maps these feature vectors to a class uses machine learning techniques. The ability to perform multi-class image classification as an automatic task using computers is increasingly becoming important. This is due to the huge volume of image data available, which are proving to be difficult for manual analysis. The difficulty arises due to of lack of human experts, poor quality images and time complexity. The current market need is to have techniques which can classify images with minimum intervention from the users in an efficient and effective manner.

Assigning images to pre-defined categories by analyzing the contents is defined as 'Image classification or 'Image categorization'. The process of image classification allows users to find desired information faster by searching only the relevant categories and not the whole information space. Image classification normally involves the process of two main tasks:-

Feature extraction task – Extracts image features and forms a feature vector and Classification task – Uses the extracted features to discriminate the classes.

CL-CL Model

Here a new approach of two level classification namely(CL-CL) is used during recognition. In order for the classification algorithms to learn to classify leaf images to a particular plant category, the training data must include both types of test cases. Let Tr and Te denote the training and testing partitions. A constant x is included to indicate the percentage of leaf feature data used for training and testing. The study analyzes different percentages of leaf feature data for this purpose and the experiments use x = 80%. That is, 80% of feature dataset is used to train level 1 and level 2 classifiers, while the rest 20% is used for testing. Now using Tr, level 1 classifier is trained.

Next, a new training feature set is constructed using only those feature samples that produce correct results. This forms a new feature set, Tr'. Tr' has the advantage of having a set of features that can improve the performance of the learning process of a classifier, as it includes only the positive results. The final step of CL-CL model, uses Tr' to identify the leaf images. As later will be proved through experiments, the refined training set improves the accuracy and reduces the error rate of the recognition process. The steps are consolidated in a diagrammatic form in Figure 4.

Group 1 : BPNN-Based CL-CL Models – Consisting of BPNN-BPNN, BPNN-SVM, BPNN-WNN models

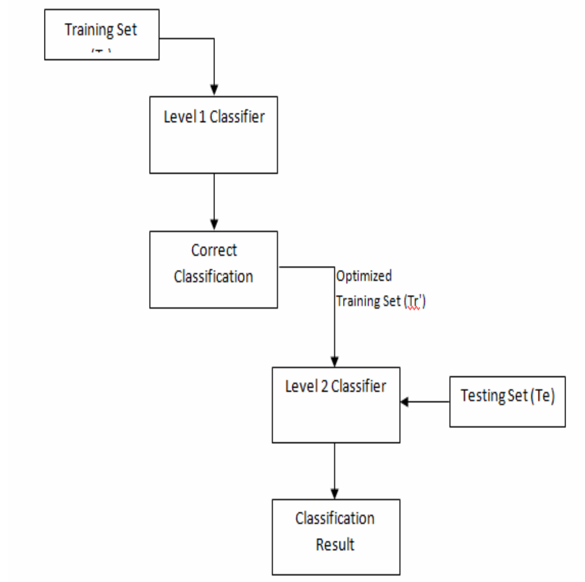

Figure 4 : Classification Process

The aim of the research is to find which of these classifiers is best suited for level 1 and which produce high performance when used for level 2 classification of CL-CL model. For this purpose, a combination of these algorithms was used to build three CL-CL hybrid models as shown in Table 1. The nine models built are grouped into three categories based on the classifier used in level 1. The three categories are :

TABLE I. PROPOSED CL-CL TWO LEVEL MODELS

I Level Classifier	II Level Classifier	Code Used
BPNN	BPNN	BPNN-BPNN
	SVM	BPNN-SVM
	WNN	BPNN-WNN

Three classifiers, namely, BPNN, SVM and WNN were used for this purpose, using which three different CL-CL models were built. The effect of the various features on recognition along with the performance of the 2-level classifier was analyzed using the two datasets, namely, real and standard, with several performance metrics were studied. The results of such experimentations are presented.

IV. RESULTS AND DISCUSSION

This section presents the segmentation results obtained by the wavelet based segmentation model and the proposed WCF Method. The evaluation is performed by analyzing the speed of segmentation and visual analysis.

To analyze the time complexity of the proposed algorithm, the time taken to segment the leaf image was calculated in seconds and the result is presented in Figure.

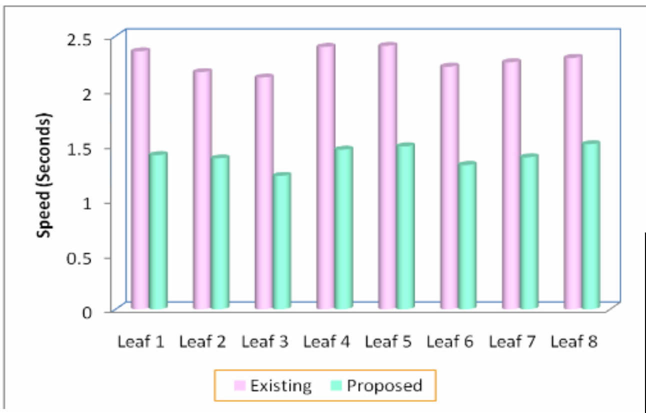

Figure 5 : Speed of Segmentation (Seconds)

The speed of the proposed algorithm is also increased considerably when compared with the existing wavelet based algorithm. On an average the existing algorithm took 2.28 seconds to segment the leaf image, while the proposed algorithm took only 1.4 seconds to extract the leaf from its background. This shows that the enhanced WCF algorithm is efficient in the process of segmentation.

BPNN-BASED CL-CL MODELS

This section presents the performance of BPNN-based CL-CL models that use BPNN classifier in level 1 and BPNN, WNN and SVM classifiers used in level 2 to produce the optimized training set. All the tables indicate the results produced only by the level 2 classifiers. The results, apart from comparing the three proposed BPNN-based CL-CL models, are also compared with the conventional BPNN classifier.

The recognition rate, error rate and speed of classification obtained by the proposed and BPNN models while using standard leaf image database are presented in this section.

Recognition Rate

Table 2 shows the recognition rate obtained by the BPNN-based CL-CL models while using the single and fused feature sets for standard leaf image database.

From the results, it is evident that all the proposed BPNN-based CL-CL models produce improved results when compared with the conventional BPNN model. Comparison between the proposed models revealed that the model that used WNN in level 1 produced the best recognition rate when compared with the other two proposed models (BPNN-BPNN and SVM- BPNN).

Among the five single feature sets, the leaf feature set followed by fractal feature set produces high recognition rate when compared with color, texture and geometric feature sets. Usage of color feature set results in lower recognition rate. This is because, by nature, the leaves consist of varying colors and usage of color features alone will not produce enough distinction in feature vector. This when combined with leaf feature vector, improves the recognition rate, proving that the leaf feature is important for leaf recognition and plant identification.

Table II RECOGNITION RATE

FEATURES	BPNN-BPNN	SVM-BPNN	WNN-BPNN	BPNN
Texture	86.41	89.85	90.34	83.52
SINGLE FEATURE SETS				

From the results of BPNN-based CL-CL models, the winning algorithm is identified as the WNN-BPNN model.

V. CONCLUSION

The field of leaf recognition for plant classification has experienced an increased need for fast and efficient classification algorithms to aid in keeping track of important plants. This chapter presented the results pertaining to the various proposed CL-CL models. The three performance metrics, namely, recognition rate, error rate and speed was used for this purpose. The proposed CL-CL model was grouped into three categories based on the classifier used in level 2. Segmentation in this area includes plant identification and plant classification. The ultimate goal of our approach is to develop a system where user can easily process a unknown plant leaf with efficient image processing and computing technique within fraction of seconds. The effectiveness of the system is proved by quantitative approach. The performance of the proposed work is compared with the existing traditional classification algorithm and real time unknown plant leaf image.

REFERENCES

1.Wu. S.G., Bao.F.S., Xu, E.Y., Wang, Y.X., Chang, F.Y. and Xiang, Q.L. (2007) Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network, IEEE International Symposium on Signal Processing and Information Technology, Pp.11 – 16.

2.Belhumeur. P.N, Chen. D, Feiner. S, Jacobs. D.W, Kress, W.J, Ling. H, Lopez. I, Ramamoorthi. R, Sameer.S, White. S and Zhan.L (2008), Searching the World's Herbaria: A system for visual identification of plant species, European Conference on Computer Vision (ECCV), Part IV, LNCS 5305, Springer-Verlag Berlin Heidelberg, Pp. 116–129.

3.Ozden. M and Polat.E, (2007), A Color Image Segmentation Approach for Content Based Image Retrieval, Pattern Recognition, Vol. 40, Pp. 1318-1325.

4.Jameel.S and Manza. R.R (2012),Color image segmentation using wavelet, International Journal of Applied Information Systems Vol.1, No.6, Pp.1-4.

5.Sengur, A. and Guo, Y. (2011), Color texture image segmentation based on neutrosophic set and wavelet transformation, Computer Vision and Image Understanding, Vol. 115, Pp. 1134-1144.

6.Priya, M and Gobu, C.K. (2013) A Wavelet based Method for Text Segmentation in Color Images, International Journal of Computer Applications, Vol. 69, No.3, Pp. 14-17.

7.Cheng, H.D., Jiang, X.H., Sun, Y. and Wang, J.L. (2001) Color image segmentation: advances and prospects, Pattern Recognition, Vol. 34, No.12, Pp. 2259–2281.

8.Comanicui, D. and Meer, P. (2002) Mean shift: a robust approach toward feature space analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No. 5.

9.Xu, B., Fu, C. and Ma, J. (2000) Image enhancement method based on wavelet transform", Proceedings of the SPIE, Vol. 4044, Pp. 150–157.

10.Du J.X., Wang, X.F., Zhang, G.J. (2007) Leaf shape based plant species recognition, Applied Mathematics and Computation, Vol. 185, Pp. 883–893.

11.Pornpanomchai, C., Rimdusit, S., Tanasap, P. and Chaiyod, C. (2011) Thai Herb Leaf Image Recognition System (THLIRS), Kasetsart J. (Nat. Sci.) , Vol.45, Pp. 551 – 562.

12.Tzionas, P., Papadakis, S.E. and Manolakis, M. (2005) Plant leaves classification based on morphological features and a fuzzy surface selection technique, Fifth International Conference on Technology and Automation, Thessaloniki, Greece,Pp. 365-370.

13.Tang, Y.Y. and Tao, Y. (1999) Feature extraction by fractal dimension, Proceedings of 5th International Conference on Document Analysis and Recognition (ICDAR '99), Pp. 217-220.

14.Li, X., wang, L. and Sung, E. (2004) Multilabel SVM active learning for image classification, International Conference on Image Processing ICIP '04, Vol. 4, Pp.2207-2210.

15.Basu, A., Watters, C., and Shepherd, M. (2002) Support Vector Machines for Text Categorization, Dalhousie University, Halifax, Nova Scotia, Canada, Proceedings of the 36th Hawaii International Conference on System Sciences, Pp.43-49.

16.Cortes, C. and Vapnik, V. (1995). Support vector networks. Machine Learning, Vol. 20, Pp.1-25.

17.Chaabane, S.B., Sayadi, M., Fnaiech, F. and Brassart, E. (2010) Colour Image Segmentation Using Homogeneity Method and Data Fusion Techniques, EURASIP Journal on Advances in Signal Processing, Vol. 2010, Article ID 367297, Pp.1-11.

18.Tang, L., Tian, L. and Steward, B.L. (2003) Classification of broadleaf and grass weeds using gabor wavelets and an artificial neural network, Transactions of the ASAE, Vol. 46, No. 4, Pp. 1247-1254.

19.Chen, J., Pappas, T.N., Mojsilovic, A. and Rogowitz, B. (2003) Image segmentation by spatially adaptive color and texture features, in: Proceedings of IEEE International Conference on Image Processing (ICIP'03), Barcelona, Spain.

20.Lee, C. and Chen, S. (2003) Classification for Leaf Images, 16th IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP 2003)Vol. 8, Pp.355-362