PLANT SPECIES IDENTIFICATION USING LEAF IMAGE AND CLASSIFICATION METHODS

Saguna Kailas Ingle[1], Vijayalakshmi Kandroli[2]

[1, 2] *Department of Information Technology, Terna College of Engineering, Navi Mumbai, (India)*

ABSTRACT

This paper is highlights plant leaf classification with the help of leaf image and its methods which can be used for plant leaf classification. A classification problem deals with associating a given input pattern with one of the distinct classes. Plant leaf classification is a technique, where leaf is classified based on its different morphological features. There are various successful classification methods like k-Nearest Neighbor Classifier, Probabilistic Neural Network and Genetic algorithm. Selection of an appropriate method for classification is often a difficult task because the quality of the results can be different for different input data. Plant leaf classifications have wide applications in various fields such as botany, Ayurveda, Agriculture etc.

Keywords: Image Processing, Leaf Classification, Plant Identification

I. INTRODUCTION

Plant recognition or classification has a broad application prospective in agriculture and medicine [1-4]. Plant leaf classification finds application in botany and in tea, cotton and other industries. Plants are vitally important for environmental protection [5]. However, it is an important and difficult task to recognize plant species on earth. Many of them carry significant information for the development of human society. So it is very necessary to set up a database for plant. The first step is to teach a computer how to classify plants.

Plants are basically identified based on flowers, fruits, and leaf [2, 6, 4]. However,fruits and flowers are three dimensional objects and increases complexity. Plant identification based on flowers and fruits require morphological features such as number of stamens in flower and number of ovaries in fruits. Identifying plants using flowers and fruits are very time consuming task.Leaves also play an important role in plant identification. Moreover, leaves can be easily found and collected everywhere at all seasons; while flowers can only be obtained at blooming season.Shape of plant leaves is one of the most important features for characterizing various plants visually. Plant leaves have two-dimensional nature and thus they are most suitable for machine processing.Before classification can be done on basis of leaf, some pre-processing is needed and most important step prior classification is feature extraction. For classification, different methods are available, namely, k-Nearest Neighbor (KNN),Probabilistic Neural Network (PNN) and Genetic algorithm (GA).

II. LEAF IMAGE CLASSIFICATION PROCESS

Leaf Image Classification process is presented in Fig.1, it consists of i. Image Acquisition, ii. Image pre-processing, and iii. Image classification.

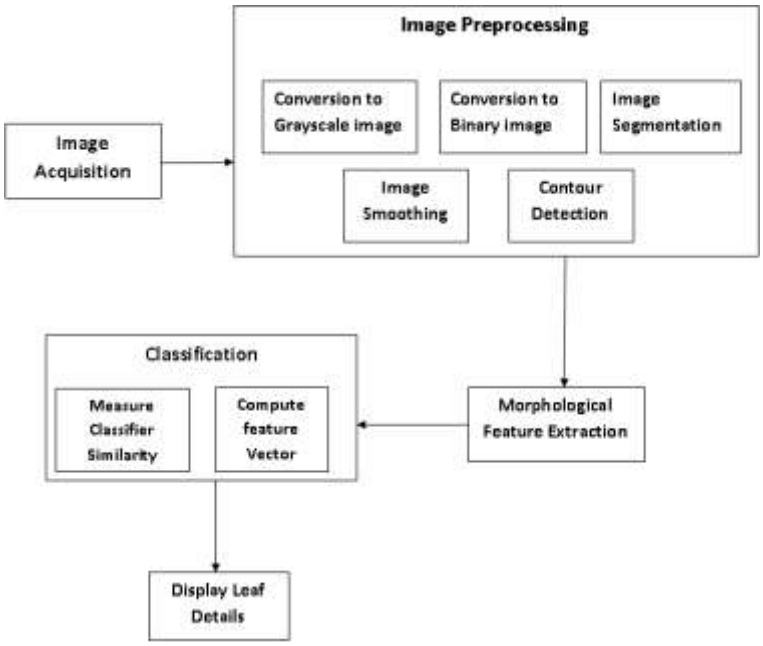

Fig 1. Leaf Image Acquisition and Preprocessing

2.1 Image Acquisition

Image acquisition includes plucking leaf from plant and then, the digital color image of the leaf is taken with a digital camera.

2.2 Pre-Processing

After leaf image is obtained some pre-processing is needed. This stage includes are (i) Grayscale conversion: a grayscale or grey-scale digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest. (ii) Image segmentation: In computer vision, image segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super-pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. (iii) Binary conversion: A binary image is a digital image that has only two possible values for each pixel. Typically the two colors used for a binary image are black and white though any two colors can be used. The color used for the object in the image is the foreground color while the rest of the image is the background color. In the document scanning industry this is often referred to as bi-tonal.Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit (0 or 1). The names black-and-white.(iv) Image smoothing: Computer vision operates on images that usually come in the form of arrays of pixel values. These values are invariably affected by noise, so it is useful to clean the images somewhat by an operation, called smoothing, that replaces each pixel by a linear combination of some of its neighbors. Smoothing reduces the effects of noise, but blurs the image. (v) Morphological features: The certain morphological features are extracted from its contour image. This feature vector is then provided to the identification methods [1,7].

III. CLASSIFICATION METHODS

3.1 K-Nearest Neighbor

In this method, each sample should be classifiedsimilarly to its surrounding samples. Therefore, if theclassification of a sample is unknown, then it could be predicted by considering the classification of its nearest neighbor samples. Given an unknown sampleand a training set, all the distances between theunknown sample and all the samples in the trainingset can be computed. The distance with the smallestvalue corresponds to the sample in the training setclosest to the unknown sample. Therefore, theunknown sample may be classified based on theclassification of this nearest neighbor[8].

3.2 Probabilistic Neural Network

Probabilistic neural networks (PNN) can be used for classification problems. PNN basically works with 3 layers. First layer is input layer. The input layer accepts an input vector. When an input is presented, first layer computes distances from the input vector to the training input vectors and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Radial Basis Layer evaluates vector distances between input vector and row weight vectors in weight matrix. These distances are scaled by Radial Basis Function nonlinearly. The last layer i.e. competitive 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.A key benefit of neural networks is that a model of the system can be built from the available data. Fig.2 shows architecture of PNN[2].

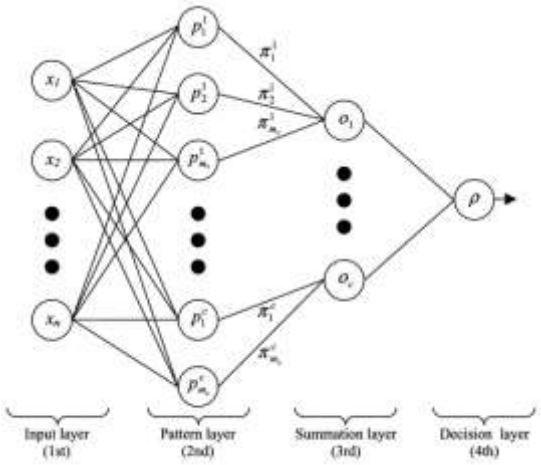

Fig. 2 Architecture of PNN

3.3 Genetic Algorithm

Genetic algorithms (GA) are inspired by Darwin's theory about evolution. Solution to a problem solved by genetic algorithms is evolved.Algorithm is started with a set of solutions (represented by chromosomes) called population. Solutions from one population are taken and used to form a new population. This is motivated by a hope, that the new population will be better than the old one. Solutions which are selected to form new solutions (offspring) are selected according to their fitness - the more suitable they are the more chances they have to reproduce.This is repeated until some condition (for example number of populations or improvement of the best solution) is satisfied[4,9].

IV. CONCLUSION

After study of all classification methods, we founddifferent advantages and disadvantages of each classification method, which are helpful for the selection of an appropriate method for a particular problem. (i) For KNN,it is a very simple method that works well on basic recognition problems but the main disadvantage of the KNN algorithm is that it is a lazy learner, i.e. it does not learn anything from the training data and simply uses the training data itself for classification and also not being robust to noisy data. (ii) PNN is much faster and more accurate,but disadvantages of PNN as it is slow and require more memory space to store the model.(iii) GA can be employed for a wide variety of optimization problems. GA performs very well for large scale optimization problems which may be very difficult or impossible to solve by other traditional methods. The disadvantage of GA is that it, sometimes; have trouble finding the exact global optimum because there is no guaranty to find best solution. Another drawback that GA requires large number of response (fitness) function evaluations depending on the number of individuals and the number of generations. Therefore, genetic algorithms may take long time to evaluate the individuals.

REFERENCES

[1] Basavaraj S. Anami, Suvarna S. Nandyal, A. Govardhan, A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal Plants, *International Journal of Computer Applications*, 6(12), 2010, 0975 – 8887.

[2] Simranjit Kaur Dhindsa, Rajbir Singh, Plant Identification and Classification Using Fuzzy Methods of Segmentation, *International Journal of Advanced Research in Computer Science and Software Engineering*,4(9), 2014, 579-582.

[3] A.H. Kulkarni, H.M.Rai, K.A. Jahagirdar, P.S. Upparamani, A Leaf Recognition Technique for Plant Classification Using RBPNN and Zernike Moment, *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1), 2013, 984-988.

[4] Anant Bhardwaj, Manpreet Kaur, A review on plant recognition and classification techniques using leaf images, *International Journal of Engineering Trends and Technology*, 4(2), 2013, 86-91.

[5] W.C. Wang,K. Freemark, The Use of Plants for Environmental Monitoring and Assessment, *Ecotoxicology and Environmental Safety*, 30(3), 1995, 289–301.

[6] James S. Cope, David Corney, Jonathan Y. Clark, Paolo Remagnino, Paul Wilkin, Plant species identification using digital morphometrics: A review, *Expert Systems with Applications*, Vol. 39(8), 2012, 7562–7573.

[7] N.Valliammal1 and S.N.Geethalakshmi, Automatic Recognition System Using Preferential Image Segmentation for Leaf and Flower Images,*Computer Science & Engineering: An International Journal*, 1(4), 2011, 13-25.

[8] Gurpreetkaur, HimanshuMonga,Classification of Biological Species Based on Leaf Architecture–A review, *International Journal of Computer Science and Information Technology & Security*, Vol. 2(2), 2012, 332-334.

[9] Valliammal Narayan and GeethalakshmiSubbarayan, An Optimal Feature Subset Selection Using GA for Leaf Classification, *The International Arab Journal of Information Technology*, 11(5), 2014, 447-451.