Literature Review of Image Features and Classifiers Used in Leaf Based Plant Recognition Through Image Analysis Approach

Amala Sabu
Department Of Computer Science
College of Engineering Poonjar
amalasabu1992@gmail.com

Sreekumar K
Department Of Computer Science
College of Engineering Poonjar
sree@cep.ac.in

Abstract—Plants play an important role in Earth's ecology by providing sustenance, shelter and maintaining a healthy atmosphere. Some of these plants have important medicinal properties. Automatic recognition of plant leaf is a challenging problem in the area of computer vision.An efficient Ayurvedic plant leaf recognition system will beneficial to many sectors of society which include medicinal field,botanic research etc. With the help of image processing and pattern recognition ,we can easily recognize the leaf images.This paper gives a survey on different leaf recognition methods and classifications. Plant leaf classification is a technique where leaf is classified based on its different features.

Keywords— ANN(Artificial Neural Network), KNN(k-nearest neighbors), PNN(Probablistic Neural Network),SVM(Support Vector Machine)

I. INTRODUCTION

Plants exist everywhere on the earth . Plant identification has become a challenging process and an active area of research. The plants are the main source of oxygen which they release by the process of photosynthesis. Apart from this, the plants are used in a variety of industrial applications such as herbs and ingredients in ayurvedic medicines, biofuels, biomass, etc[13]. India has vast history of using plants as a source of medicine. This science is termed as Ayurveda[1]. According to Ayurveda every plant on earth has some medicinal value. It is considered a form of alternative to allopathic medicine in the world. One of the major advantages is that it does not have any side effects. These medicinal plants are identified manually by taxonomist,which are prone to human errors in many cases. To avoid this, develop an automated system for plant leaf recognition. Many researchers have made an attempt for medicinal plant identification using the images of leaf. Researchers are using different techniques for identification. Some researchers are using shape features while others are using texture features. Also they using different classifiers. Leaf recognition is generally based on the observation of the morphological characteristics of leaf, but it will be a difficult task for experienced botanists to identify the plants because of the large number of leaves existing in the world. In this case, it is helpful for developing an efficient plant recognition system based on computer to identify the leaves. With the development of image processing it is available to apply them to recognize plant automatically. Many studies in the past have shown that leaf contains rich information like color, shape etc. In order to identify plants that are unfamiliar and to differentiate the medicinal plants with similar features, it is necessary to classify them with distinguishing features. The large variety in medicinal plants are identified by features, which are also used for grouping them. We used physical characteristics called morphological features for identification. The shape of a leaf is an important feature. The shape of leaves often vary from species to species. Classify the plants into their appropriate species using classifiers like SVM, KNN,ANN etc.

II. LITERATURE SURVEY

Plants plays a considerable part in both human life and other lives that is present on the earth. Plant recognition depending on images of leaf. Some systems employ descriptions used by botanists . But it is difficult to extract and transfer those features to a computer automatically. In the present day scenario, the ready availability of digital cameras and handheld computers has made the possibility of the development of this system real. The studies such as image processing and machine learning have helped the researchers to work for nonmanual classification of plants. The challenge in developing this system is the identification of discriminative features for various species. The properties which are common for a group and uncommon in other groups

978-1-5090-5297-4/17/$31.00 ©2017 IEEE

are identified through feature extraction. Basic step for leaf recognition shown below.

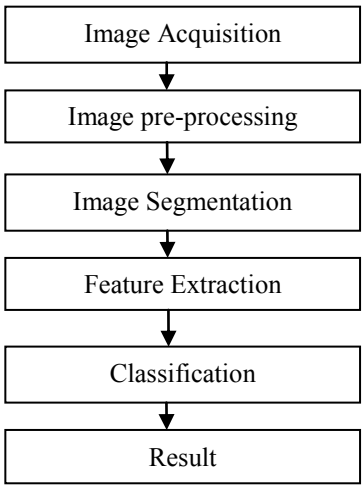

Fig.1 Basic steps for leaf recognition

The images of the plant leaf are captured through the camera. To remove noise in image or other object removal, different pre-processing techniques is considered (Eg.Image clipping, Image smoothing). Segmentation means partitioning of image into various parts of same features or having some similarity. Feature extraction plays an important role for identification of an object. In many application of image processing feature extraction is used. Morphology, color, texture, edges etc. are the features used for leaf recognition. After feature extraction is done, the database images are classified . From these methods, we can accurately identify the leaves.

A. REVIEW OF PRE-PROCESSING

The objective of the pre-processing step is to standardize the scale and orientation of the image before feature computation. The raw image is typically a color image oriented at a random angle and having a random size. The image is first converted to binary and gray scale forms. To make features rotation-invariant, the angle of the major axis of the leaf is extracted from the image and used to rotate it so that the major axis is aligned with the horizontal line[14].

- Gray scale conversion : The image converting to gray scale. The gray scaled images were subjected to the process of image contrast and intensity enhancement techniques and then stacked together as slices for further processing.
- Binary conversion : Thresholding can be used to create binary images from a gray scale image.Binary image is a digital image that has only two possible values for each pixel. Commonly, the two colors used for a binary image are black and white.
- Noise Removal : Digital images are prone to a variety of types of noise. Noise is the result of errors in the digital image acquisition process that result in pixel values We can use linear filtering to remove certain types of noise. Certain filters, such as averaging,low pass filter or Gaussian filters, are appropriate for this purpose. For example, an averaging filter is useful for removing grain noise from a photograph. Salt and pepper noise from an image remove using an averaging filter and a median filter .
- Contrast stretching : Also called Normalization.Some images are homogeneous, they do not have much change in their levels. In terms of histogram representation,they are characterized as the occurrence of very narrow peaks. The homogeneity can also be due to the incorrect illumination of the scene. Ultimately the images hence obtained are not easily interpretable due to poor human perceptibility. This is because there exists only a narrow range of gray-levels in the image having provision for wider range of gray-levels.The contrast stretching methods are designed exclusively for frequently encountered situations. Different stretching techniques have been developed to stretch the narrow range to the whole of the available dynamic range.
- Histogram Modification : Histogram has a lot of importance in image enhancement. It reflects the characteristics of image. By modifying the histogram, image characteristics can be modified. One such example is Histogram Equalization. Histogram equalization is a nonlinear stretch that redistributes pixel values so that there is approximately the same number of pixels with each value within a range. The result approximates a flat histogram. Therefore, contrast is increased at the peaks and lessened at the tails.

B. REVIEW OF FEATURES

Shape Features

- Length of leaf **:** This is the distance between the two ends of the main vein of the leaf[1].
- Breadth of leaf **:** This is the distance from the left most point in a leaf to the right most point in the leaf.
- Aspect Ratio **:** The aspect ratio of a leaf is the ratio of the length to its breadth.
- Diameter **:** The diameter of a leaf is the maximum distance between any two points which lie inside the area covered by the leaf.
- Shape of leaf (Convex Hull) **:** Using the Convex Hull algorithm we can find the coordinates of the points under which the entire area of the leaf is covered
- Leaf perimeter **:** Leaf Perimeter is calculated by counting the number of pixels consisting leaf margin.
- Rectangularity (R) **:** The process includes drawing a

rectangle outside the image, so that the image just fits in the rectangle. R= LpWp/A, where Lp is length,A is leaf area and Wp is breadth.

- Leaf Area **:** Smoothed leaf image is consider to find out leaf area. Number of pixels having binary value 1 is termed as leaf area[10].
- Compactness : It is defined as the ratio of the product of area with 4π to the square of perimeter .It is also referred to as roundness[2]
- Eccentricity :It is a characteristic feature of any conic section ,where 'b' and' a 'refer to minimum and maximum axial length respectively for an ellipse.

$$Eccentricity = \sqrt{1-(b/a)^2}$$

Texture Features

- Energy:

$$E = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(P(\mathrm{i},\mathrm{j}))^2$$

- Contrast :

$$C = \sum_{n=0}^{G-1} n^2 \{\sum_{i=1}^{G}\sum_{j=1}^{G} P(\mathrm{i},\mathrm{j})\} \;;\; |\mathrm{i}-\mathrm{j}|=\mathrm{n}$$

- Entropy **:**.

$$H = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(P(\mathrm{i},\mathrm{j})\log P(\mathrm{i},\mathrm{j}))$$

Where G is the number of gray levels,P(i,j) is the probability distribution in the GLCM[7].

Color Features

- Arithematic Mean :

$$mean = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} P(\mathrm{i},\mathrm{j})$$

- Standard Deviation :

$$SD = \left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(\mathrm{P}(\mathrm{i},\mathrm{j})-\mathrm{mean})\right]^{1/2}$$

- Skewness(Θ) :

$$\Theta = \frac{1}{MN\sigma^3}\sum_{i=1}^{M}\sum_{j=1}^{N}(\mathrm{P}(\mathrm{i},\mathrm{j})-\mathrm{mean})^3$$

- Kurtosis :

$$\Upsilon = \frac{1}{MN\sigma^4}\sum_{i=1}^{M}\sum_{j=1}^{N}(\mathrm{P}(\mathrm{i},\mathrm{j})-\mathrm{mean})^4$$

Where M and N are the dimensions of the image,P(i,j) are values of the color on column i and row j[8].

Other Features

- Speeded Up Robust Features(SURF) **:**

  It is a local feature detector and descriptor, it is used for object recognition and classification . SURF is based on sums of 2D Haar wavelet responses [4].

- Scale-invariant Feature Transform (SIFT) **:**

  Used to detect and describe local features in images. The Scale-Invariant Feature Transform (SIFT) combines together a feature descriptor and a detector. The feature detector as the name itself suggest basically extracts from an image a number of frames in a way which is consistent despite of few variations of the illumination ,viewpoint and other conditions of viewing. The descriptor associates with the regions a signature which basically identifies their appearance robustly and compactly[3].

- Histogram of Oriented Gradients(HoG) **:**

  The hog feature descriptors used for the purpose of object detection which counts orientation in localized portions of an image. This method is similar to that of scale-invariant feature, but differs in that it is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy[5].

C. REVIEW OF CLASSIFIERS

- *K-Nearest Neighbor(KNN)*

k-nearest neighbor algorithm is a method for classifying objects based on closest training examples in the feature space. k-nearest neighbor algorithm is among the simplest of all machine learning algorithms. Training process for this algorithm only consists of storing feature vectors and labels of the training images. In the classification process, the unlabelled query point is simply assigned to the label of its k nearest neighbors. Typically the object is classified based on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be a odd integer.

However, there can still be ties when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean distance[6].

- Probabilistic Neural Network (PNN)

A probabilistic neural network **(**PNN**)** is a feed forward neural network . PNN is often used in classification problems.When an input is present, the first layer computes the distance from the input vector to the training input vectors. This produces a vector where its elements indicate how close the input is to the training input. The second layer sums the contribution for each class of inputs and produces its net output as a vector of probabilities. Finally, a compete transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 (positive identification) for that class and a 0 (negative identification) for non-targeted classes..In a PNN,the operations are organized into a multilayered feed forward network with four layers.Four layers are Input layer, Hidden layer, Pattern layer/Summation layer and Output layer[9].

- Support Vector Machine(SVM)

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships[12]. An SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. It can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier[11].

- Decision Tree Classifier

Decision tree is an approach for predictive modeling used in data mining and machine learning, by which the data in data sets can be classified into classes. By using decision trees, the system eliminates half of the cases at each step. This is extremely beneficial because in a system like this, where the dataset is huge in number, the processing takes a chink of the total time. Hence, it has to be minimized wherever possible. The decision tree classifiers organized a series of test questions in a tree structure. The final result is a tree with decision nodes and leaf nodes. Each internal node represents a "test" on an attribute. Each branch represents the outcome of the test. Follow the appropriate branch based on the outcome of the test. It then leads us either to another internal node or to a leaf node[1].

- Artificial Neural Network (ANN)

Artificial Neural Network have proven themselves as proficient classifiers and are particularly well suited for addressing non-linear problems like leaf classification. ANN is an interconnected group of nodes. Neural networks consist of multiple layers and the signal path traverses from front to back. Back propagation is where the forward stimulation is used to reset weights on the "front" neural units and this is sometimes done in combination with training where the correct result is known[7].

III. COMPARATIVE STUDY

Classification Techniques	**Advantages**	**Disadvantages**
KNN Classifier	1. Simplest 2.Robust with regard to search space 3.No training is required, confidence level can be obtained	1. Expensive testing of each instance 2. Sensitiveness to noisy or irrelevant inputs 3.Lazy Learning
Probabilistic Neural Network	1. Tolerant of noisy inputs 2. Instances can be classified by more than one output 3. Adaptive to changing data	1. Long training time 2. Large complexity of network structure 3. too many attributes can result in over fitting
Support Vector Machine	1. Good generalization capability 2. Sparseness of the solution and the capacity control obtained by optimizing the margin 3. SVMs can be robust, even when the training sample has some bias	1. Slow training 2. Difficult to understand structure of algorithm 3. limitations in speed and size, both in training and testing
Decision Tree	1.Decision trees are simple to use, easy to understand 2.System eliminates half of the cases at each step 3. Fast prediction	1.A small change in input data can at times, cause large changes in the tree.
Artificial Neural Network	1.Requiring less formal statistical training 2.Ability to implicitly detect complex nonlinear relationships between dependent and independent variables	1. Greater computational burden 2. Proneness to overfitting

Table 1: Comparison table

IV. CONCLUSION

There are many techniques relevant for the purpose automated recognition for plant leaf recognition. This survey is used to identify different features and classifiers which are used for medicinal plant leaf recognition. From study of above classification techniques we come up with a comparison table. The nearest-neighbor method is the simplest classification technique.The disadvantage of the KNN method is the time complexity of making predictions.

V. ACKNOWLEDGMENT

This study would not have been possible without the support of many people. First and the foremost, I give thanks to Almighty God who gave me the inner strength, resource and ability to complete my study. I would like to express my sincere thanks to Mr. Sreekumar K, Assistant Professor,CEP for his valuable help and support.

References

[1] D Venkataraman, Siddharth Narasimhan, Shankar N, S Varun Sidharth,Hari Prasath D ”Leaf Recognition Algorithm for RetrievingMedicinal Information”,Springer International Publishing AG 2016

[2] A Gopal, S. Prudhveeswar Reddy “Classification of Selected Medicinal Plants Leaf Using Image Processing” -IEEE,2012

[3] Shubham Lavania,“Leaf Recognition using Contour based Edge Detection and SIFT Algorithm”- IEEE International Conference,2014

[4] Qunang-Khue Ngueyn “Leaf based plant identification system for antroid using SURF features in combination with bag of words model and supervised learning”-,International Conference on Advanced Technologies for communicatuons 2013

[5] Xue-Yang Xiao, Rongxiang Hu “HOG-Based Approach for Leaf Classification”- Springer International Publishing 2010

[6] Snehal Shejwal,Prachi Nikale,Ashwini Datir “Automatic Plant Leaf Classification on Mobile Field Guide”- International Journal of Computer Science And Technology,2015

[7] R Janani,A Gopal “Identification of selected Medicinal Plant Leaves Using Image Fatures and ANN”-International Conference on Advanced Electronic Systems 2013

[8] Pushpa BR, Anand C, Mithun Nambiar P “Ayurvedic Plant Species Recognition using Statistical Parameters on Leaf Images” - International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 7 (2016)

[9] Shreepad S. Sawant, Preeti S. Topannavar “Introduction to Probabilistic Neural Network –Used For Image Classifications”- International Journal of Advanced Research in Computer Science and Software Engineering, 2015, Volume 5

[10] Sandeep Kumar.E “Leaf color, area and edge features based approach for identification of Indian medicinal plants”- Indian Journal of Computer Science and Engineering (IJCSE).

[11] Shivaputra S.Panchal, Rutuja Sonar “Pomegranate Leaf Disease Detection Using Support Vector Machine”- International Journal Of Engineering And Computer Science,2016

[12] Sethulekshmi A V, Sreekumar K ”Ayurvedic leaf recognition for Plant Classification”- International Journal of Computer Science and Information Technologies, Vol. 5 (6) , 2014

[13] Divya Tomar and Sonali Agarwal” Leaf Recognition for Plant Classification Using Direct Acyclic Graph Based Multi-Class Least Squares Twin Support Vector Machine” - International Journal of Image ,2016

[14] Jyotismita Chaki, Ranjan Parekh and Samar Bhattacharya “Plant Leaf Recognition Using Ridge Filter and Curvelet Transform with Neuro-Fuzzy Classifier”- Springer International Publishing 2016