International Journal of Advances in Engineering & Technology, May, 2014.
©IJAET ISSN: 22311963

A SURVEY ON PLANT SPECIES CLASSIFICATION BASED ON MORPHOLOGICAL FEATURES OF LEAF

Vandana C. Bagal[1] and Ramesh R. Manza[2]
[1]Department of MCA Engineering, KKWIEER, Pune University, Nashik, India
[2]Department of Computer Science, Dr. Babasaheb Ambedkar University, Aurangabad, India

ABSTRACT
Plants are very important to us. All food people eat comes directly or indirectly from plants. To effectively use plants, one must learn basic plant identification and classification skills. So it is necessary to establish database by information technology as soon as possible. The distinct feature of plant and their leaf motivate the researchers to perform research on it to accurately classify the plants. In this paper we present overview of various feature extraction technique and plant classification methods which can be useful for accurately classifying the plants. We review the computational morphometric that have been used in recent years to analyse images of plant leaf. In this paper we have discussed various plant species classification methods where leaf is classified based on its digital morphological features. Among the available methods of classification deciding the best one is really a difficult task because accuracy of the result is depend on the input data. The aim of this survey is to provide an overview of different plant leaf feature extraction and plant classification technique.

KEYWORDS: *Morphological feature, plant leaf classification, feature extraction, moment feature.*

I. INTRODUCTION

In order to recognize plant species from their shapes, color, textures and seedling, there have been many recent studies on plant classification and recognition. It is very difficult to analyze the shape of flowers and seedling because of their 3D structure if we consider only 2D images. So in this survey we have concentrated on plant species classification based on morphological features of leaf. Taxonomy relies greatly on morphology to discriminate groups. Computerized geometric morphometric methods for quantitative shape analysis measure, test and visualize differences in form in a highly effective, reproducible, accurate and statistically powerful way. Plant leaves are commonly used in taxonomic analyses and are particularly suitable to landmark based geometric morphometrics. Plant are useful for environmental protection. Due to deterioration of environment like global warming large number of plant species are at the margin of extinction. Many of rare plant species have died out. So it is necessary and urgent need to classifying plant species by its category to help botanist by establishing bio-diversity database. The objective of this research paper is plant species classification based on features of leaf. The organization of this paper is section 1 describes introductory part , section 2 shows a brief literature review , section 3 represents various feature extraction techniques, various methods of classification represents section 4. Finally section 5 concludes this paper.

II. PREVIOUS WORK

Plant play an important role in both human being life and other lives that are exist on the universe. Studies have been done on the automation of plant classification and recognition. A large percentage of such works were based on the extraction of single feature from the image of a plant part such as the leaf, or the flower [1]. Plant species identification based on leaf, fruit and flower is plays important role in the domain of pattern recognition. Plant Recognition based of images of leaf is a very challenging task. An approach for automatically identifying plant species from photos of the leaves is presented by Stefen Fiel [2]. In this first the images are transformed into grayscale and normalized. This paper proposed to use local features to avoid the segmentation step and to overcome the effect of damaged leaves, bad segmentation and occlusions to the shape features. Additional a bag-of-key point's method is used and the classification is done with a one-vs-all support vector machine. [3] In this paper, a digital morphological feature based automatic recognition method for plant images was proposed and performed. The fifteen features are used to classify 20 species of plant leaves. In addition a new move median centers hypersphere classifier is adopted to perform the classification. A novel approach that integrates a thresholding technique and H-maxima transformation based technique is developed by Valliammal and Geethalakxhmi[4] to extract the leaf veins. A preliminary segmentation depending on the intensity histogram of leaf images is carried out initially to roughly determine vein regions by means of thresholding. This is followed by means of H maxima transformation based technique for object pixel as its outputs. A novel approach of gabor filter banks specifically designed for plant species recognition by using their bark texture characteristics. In this paper by Zheru Chi et al[5] developed , a texture is modeled as numerous narrowband signal that are distinguished by their central frequencies and normalized ratios of amplitudes. The normalized ratio of amplitudes is utilized as an energy weight for integrating narrowband signals. In accordance with this texture model, a collection of texture feature can be obtained from each kind of plant bark that is explored to differentiate the plant and to design the equivalent Gabor filter bank. A classifier is built by these gabor filter banks. A paper by Brendon J. Woodford et al[7]. In this paper they have used wavelet based analysis of fruit/leaf images for doing classification. The fruit/leaf images consist of the fruit and insect which is damaging it; therefore it is a classification problem which tries to identify fruit/leaf having particular pests damaging it. A paper by S. G. Wu, F.S. Bu[6] PNN (Probabilistic Neural Network) has been used for plant species classification based on leaf architecture. They have extracted 12 features of the leaf which are further orthogonal into five principle variables which consist of the input vectors. PNN has been trained on very large number of leaves and has classified 32 species of plants. This paper gives accuracy 90%. An uncomplicated and computationally effective technique for plant species recognition by means of leaf image is recommended by Hossain and Amin [9]. This technique executes only on the plants with wide flat leaves which are more or less two dimensional in general. This technique includes five major phases. In first phase, images of leaf are obtained with digital camera or scanners. Subsequently the user decides the base point of the leaf and a small number of reference points on the leaf blades. In accordance with these points the leaf shape is obtained from the background and a binary image is generated. Following that the leaf is aligned horizontally by keeping its base point on the left of the image. Then a number of morphological characteristics, for instance eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area and extent are obtained. A distinctive collection of features are obtained from the leaves by segmenting across the major axis and parallel to the minor axis. After that the feature points are normalized by considering the ratio of the slice lengths and leaf lengths. These features are provided as inputs to the probabilistic neural network [13]. A paper by C. S. Sumathi has used a feed forward neural network to automate the leaf recognition for plant classification. Correlation feature selection is used for selecting features. The extracted features were trained using 10 fold cross validation and tested with neural network.[16]. Vishakha Metre has discussed the different ways in which the problem of accurate plant leaf classification has been formulated. Author has conclude that combined classifier is beast for plant images classification based on textural features.

III. FEATURE EXTRACTION TECHNIQUES

The technique to extract digital morphological feature is stated by Ji-Xian Du, Xiao-Feng Wang and Guo-Jun Zhang. The features are extracted from the contours of leaf. The digital morphology features (DMF) generally include geometrical features (GF) and invariable moment features (MF). The geometrical features consist of aspect ratio, rectangularity, area ratio of convexity, perimeter ratio of convexity, sphericity, circularity, eccentricity and form factor etc[3]. As stated by kadir et al [8] features of leaf are divided into two categories. These categories are general visual features and domain related visual features. The general visual features are consisting of color, texture and shape. Domain related visual features are integrated with morphology characteristics of a leaf are shape, dent and vein. Further Hossain and Amin [9], who stated in their publication there are several morphological features which are properties from the shape feature such as eccentricity, area, perimeter, major axis, minor axis, equivalent diameter, convex area and extent. Valliammal and Geethalaxmi [4] were confirmed in their research paper that leaf image could be categorized based on color, texture, shape or combination of these properties. Shabanzade et al[10] used statistical moments and histogram based features method in order to extract a leaf texture feature. The reason method being used because to avoid lose some significant information regarding pixel, pixel position and information of the texture. All extracted features are classified into local descriptors category. Besides thresholding and segmentation method has been used where the image has converted into binary image for separating the background and the object. Finally the extracted information has been classified into global features category. So global feature and local descriptor are two categories for features of leaf. The shape, color and skeleton are basic features for plant classification according to Jing et al [11].

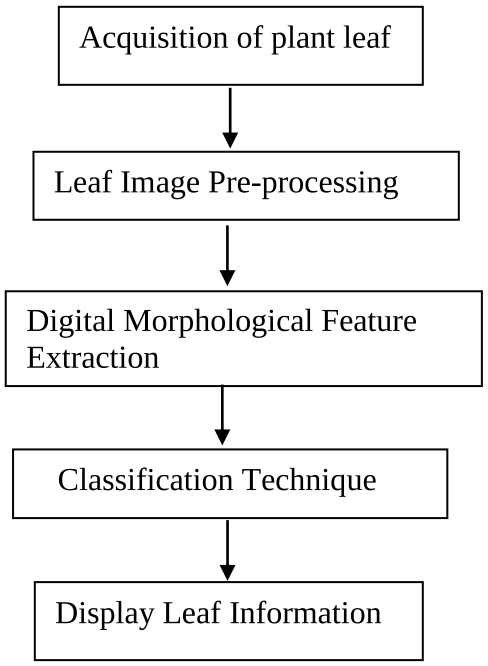

Fig.1 Block Diagram of the System

Researchers have used several digital morphological features to classify plants. Below table shows comparative study of feature extraction techniques

Tabel 1. Feature extraction techniques

Sr. No.	Research Paper	Features of Leaf	Sub Features	Data set Size
1.	Leaf Shape based Plant Species Recognition[3]	Digital Morphological features, Geometrical features, Invariable moment features	Geometrical features include -aspect ratio - rectangularity - areal ratio of convexity -perimeter ratio of convexity -perimeter ratio of convexity -sphericity -circularity -Eccentricity -form factor	20
2.	A Comparative Experiment of Several Shape Methods in Recognizing Plants[14]	General visual features Domain related visual features	General visual features include- - Color - Texture - Shape	52

			Domain related visual features - Shape - Dent	
3.	Leaf Shape Identification based Plant Biometrics[9]	Shape	Shape properties - Eccentricity - Area - Perimeter - Major axis - Minor axis - Equivalent diameter - Convex area - Extent	30
4.	Hybrid Image Segmentation Algorithm for Leaf Recognition and Characterization	Color Texture Shape	Shape features - Surface area - Surface perimeter - Disfigurement Color features - Red - Green - Blue Texture features - Texture entropy - Texture contrast - Texture energy	32
5.	Combination of Local Descriptors and Global Features for Leaf Recognition[10]	Global features Local Descriptors	Global Features - Length - Leaf area - Width Local Descriptors - Contrast - Correlation - Homogeneity	20
6.	A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network.[6]	Digital morphological features derived from 5 basic features	Basic Features – Diameter, physiological length, physiological width, leaf area, leaf perimeter Digital Morphological Features- Smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of physiological length and physiological width, vein features	32
7.	Classification of Biological Species based on Leaf Architecture	Isometric Quotient, Eccentricity, Aspect Ratio, Leaf Area, Leaf Perimeter, Area Ratio of Perimeter, Solidity, Length of Major and Minor Axis, Upper and Lower Triangle Area , Boundary Feature	-	19

IV. METHODS OF CLASSIFICATION

Various Classification methods are implemented for achieving more accuracy in identification of the plants based on features of leaf.

4.1. MMC Hypersphere Classifier

The fundamental idea of the MMC is that each class of patterns are regarded as a serried of "hyperspheres", which is treated as a set of "points" in convential approaches. The first step of MMC is to compute the multidimensional median of the points of the considered class, set the initial center as the point that is closest to that median, find the maximum radius that can encompass the points of the class. Through certain iteration we remove the center of the hypersphere in a way that would enlarge the hyperspehere and have it to encompass as many points as possible. This is performed by having the center "hop" from one data point to a neighbouring point. Once we find the largest possible hypersphere, the points inside this hypersphere are removed, and the whole procedure is repeated for the remaining points of the class. We continue until all points of that class are covered by some hyperspheres. At that point we tackle the points of the next class in a similar manner. The final step is to remove redundant hypersphere, which are totally encompassed by a larger hypersphere [3].

4.2. Probabilistic Neural Network

Probabilistic neural networks can be used for classification problems. It has parallel distributed processor that has a natural tendency for storing experimental knowledge. PNN is derived from Radial Basis Function (RBF) Network which is an ANN using RBF. PNN has three layers: the input layer, Radial Basis Layer and the Competitive Layer. Radial basis layer evaluates vector in weight matrix. These distances are scaled by Radial Basis Function nonlinearly. Then the competitive Layer finds the shortest distance among them and thus finds the training pattern closest to the input pattern based on their distance. A key benefit of neural networks is that a model of the system can be built from the available data. PNN is adopted for it has many advantages. Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise.

4.3. K- Nearest Neighbor Classifier

K Nearest neighbour classifier calculates the minimum distance of a given point with other points to determine its class. Suppose we have some training objects whose attribute vectors are given and some unknown object w is to be categorized. Now we should decide to which class object w belongs. Let us take an example. According to the K-NN rule suppose we first select k=5 neighbour of w. Because three of these five neighbours belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the K-NN rule. It is intuitive that the K-NN rule doesn't take the fact that different neighbours may give different evidences into consideration. Actually, it is reasonable to assume that objects which are close together (according to some appropriate metric) will belong to the same category. According to the K-NN rule suppose we first select k=5 neighbours of w. Because three of these five neighbours belong to class 2 and two of them to class 3, the object w should belong to class 2, according to the K-NN rule. For plant leaf classification we find out feature vector of test sample and then calculate Euclidean distance between test sample and training sample. This way it finds out similarity measures and accordingly finds out class for test sample. The k-nearest neighbour's algorithm is amongst the simplest of all machine learning algorithms. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours. K is a positive integer, typically small. If k=1, then the object is simply assigned to the class of its nearest neighbour. In binary classification problems, it is helpful to choose k to be an odd number as this avoids tied votes. It is intuitive that the k-NN rule doesn't take the fact that different neighbours may give different evidences into consideration. Actually, it is reasonable to assume that object which is close together (according to some appropriate metric) will belong to the same category.

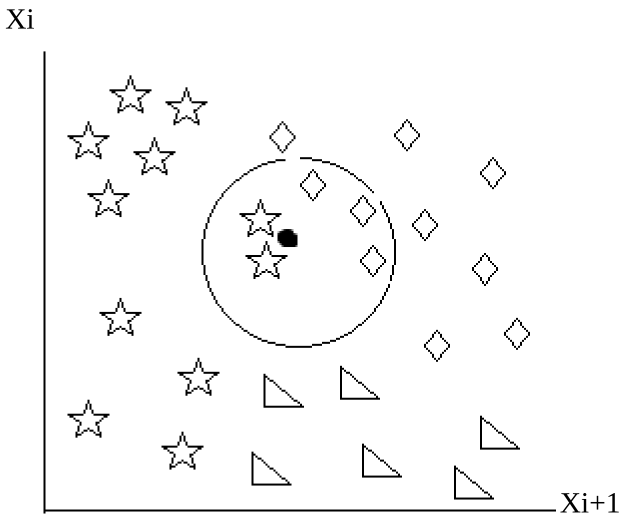

Figure 1. Example for K-nearest neighbour Classifier

4.4. K-Means Clustering

The K-means clustering algorithm is simple method for estimating the mean (vectors) of a set of K-groups. K-means clustering is a supervised learning algorithm and utilized a prior knowledge of the number of clusters. Biva [8] introduced K-means clustering for the purpose of leaf image classification. The simplicity of the algorithm also can lead to some bad solutions. Initial cluster seeds are chosen (at random). These represent the "temporary" means of the clusters. Imagine our random numbers were 60 for group 1 and 70 for group. The squared Euclidean distance from each object to each cluster is computed, and each object is assigned to the closest cluster. For each cluster, the new centroids computed and each seed value is now replaced by the respective cluster centroid. The squared Euclidean distance from an object to each cluster with the smallest squared Euclidean distance. The cluster centroids are recalculated based on the new membership assignment. The process is repeated until no object moves clusters.

4.5. Genetic Algorithm

Genetic algorithms are mainly used for feature classification and feature selection. The basic purpose of genetic algorithm is optimization. GA s give a heuristic way of searching the input space for optimal x that approximates brute force without enumerating all the elements and therefore bypasses performances issues specific to exhaustive search. Genetic algorithm is used effectively in the evolution to find a near-optimal set of connection weights globally without computing gradient information and without weight connections initialization. Though solution found by genetic algorithm is not always best solution. It finds "good" solution always. Main advantage of GA is that is adaptable and it possess inherent parallelism. Genetic algorithms handle large, complex, non-differentiable and multi model spaces for image classification and many other real world applications.

4.6. General Regression Neural Networks

General regression neural networks perform regression where the target variable is continuous. The basic idea is that a predicted target value of an item is likely to be the same as other items that have close values of the predictor variables. Jyotismita[15] used GRNN for classification scheme.

4.7. Support Vector Machine

Support vector machine is a non-linear classifier. The idea behind the method is to non-linearly map the input data can be linearly separated, thus providing great classification performance. Support vector machine is a machine learning tool and has emerged as a powerful technique for learning from data and in particular for solving binary classification problems.[12] . The main concept of SVM are to first transform input data into a higher dimensional space by means of a kernel function and then construct an OSH(Optimal Separating Hyper Plane) between the two classes in the transformed space[12]. For plant leaf classification it will transform feature vector extracted from leaf's contour.

SVM finds the OSH by maximizing the margin between the classes. Data vectors nearest to the constructed line in the transformed space are called the support vectors. The SVM estimates a function for classifying data into two classes. Using a nonlinear transformation that depends on a regularization parameter, the input vectors are placed into a high dimensional feature space, where a linear separation is employed. Main advantage of SVM is it has a simple geometric interpretation and gives a sparse solution. Unlike neural networks, the computational complexity of SVM does not depend on the dimensionality of the input space one of the bottlenecks of the SVM is the large number o support vectors used from the training set to perform classification tasks.

V. CONCLUSION

In this survey, we have discussed a brief overview of leaf feature extraction techniques for plant species classification. To the best of our knowledge we think that leaf color changes according to atmospheric condition and nutritional value, so it has less reliability for feature extraction. Seed and flowers may give correct result if one will consider 3D image of it. In this paper we have also discussed various plant species classification methods which helps to analyse the best classification method. The use of computational morphometrics helps to analyse leaf images. We believe that digital morphometrics help to define species more effectively and providing rapid species identification. So we conclude that for a well-developed leaf, digital morphological features will give best result for feature extraction.

VI. FUTURE SCOPE

According to literature survey there are various classification methods available. There is no research been done on comparative analysis of plant classification algorithm. For future scope we are planning that a comparative analysis of classification algorithm like move median centre, k nearest neighbour, probabilistic neural network, genetic algorithm etc. should be done. So in this way one can find the classification algorithm which gives more accurate result and helps the botanist in identifying plants.

ACKNOWLEDGEMENTS

I would like to thanks my parents and my husband for showing trust and giving their support which helps me successful in completing this paper.

REFERENCES

[1] Gurpreet Kaur, Himanshu Monga, April 2012 "Classification of Biological Species Based on Leaf Architecture-A review" Vol. 2, No. 2, ISSN: 2249-9555

[2] Stefan Fiel and Robert Stablanig: "Leaf Classification using Local Features"

[3] Ji-Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang:" Leaf Shape based Plant Species Recognition" Applied Mathematics and Computation 185(2007)883-893.

[4] N.Valiammal, Dr.S.N.Geethalakshmi: "Hybrid Image Segmentation Algorithm for Leaf Recognition and Charaterization", 2011 IEEE

[5]Zheru Chi, Li Houqiang and Wang Chao :" Plant Species Recognition based on Bark Patterm using Novel Gabor Filter Banks", Proceeding of International Conference on Neural Network and Signal Processing, Vol 2. Pp: 1035-1038, 2003

[6]S.G.Wu, F.S.Bao, E.Y.Xu, YU-Xuan Wang, Yi-Fan Chang and Qiao Liang Xiang : "A Leaf Recognition Algorithm for Plant Classification using Probabilistic Neural Network. In Proc. IEEE International Symposium on Signal Processing and Information Technology, Pages 11-16, 2007

[7]Brendon J. Woodford, Nikola K.Kasabov and C. Howard Wearing: "Fruit Image Analysis using Wavelets.

[8] Abdul Kadir, Lukito Edi, Nugroho and Paulus Insap Santosa:" Leaf Classification using Shape, Color and Texture", International Journal of Computer Trends and Technology (IJCTT), July-August 2011, pp 225-230.

[9]J. Hossain and M. A. Amin: "Leaf Shape Identification based Plant Biometrics", 13th Internatinal Conference on Computer and Information Technology (ICCIT) pp 458-463, 2010

[10]Shabanzade, M. M. Zahedi AND S.A.Aghvami: "Combination of Local Descriptors and Global Features for Leaf Recognition", Signal and Image Processing, Int.J. 2: 23-21, DOI: 10.5121,2011

[11]Jing, L. Z. Shanwen and L. Jiandu: "A method of Plant Leaf Recognition based on Localy Linear Embedding and Moving Center Hypersphere Classifier" Proceedings of the Intelligent Computing Technology and Application, Springer Verlag (ASV' 09), Uslan, Sauth Korea, PP: 645-651. DOI : 10.1007/978-3-642-04020-769

[12]Krishna Singh, Dr. Indra Gupta and Dr. Sangeeta Gupta, Retrieval and Classification Leaf Shape by Support Vector Machine using Binary Decision Tree, Probabilistic Neural Network and Generic Fourier moment Technique : a Comparitive
Study IADIS International Conference Computer Graphics, Visualization, Computer Vision and Image Processing 2010.

[13] Arunpriya C. Antony Selvadoss Thanamani:"A survey on Species Recognition System for Plant Classification "Int. J.Computer Technology and Applications, Vol 3(3), 1132-1135

[14] A.Kadir, LE.Nugoho, A.Susanto and P.I.Santosa: "A Comparative Experiment of Several Shape Methods in Recognizing Plants", IJCSIT, Vol. 3, No.3, June 2011

[15] Jyotismita Chaki, Ranjan Parekh: "Plant Leaf Recognition using Shape based Features and Network Classifiers (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.2, No. 10, 2011.

[16] C.S.Sumathi and A.V.Senthil Kumar : " Plant Leaf Classification Using Soft Computing Techniques" Int. J. Future Computer and Communicatin, Vol.2, No.3, June 2013.

[17] Vishakha Metre and Jayashree Ghorpade : " An overview of the Research on Texture Based Plant Leaf Classification: Int. J. of Computer Science and Network, Vol 2, Issue 3,2013.

AUTHORS

V. C. Bagal received the BSc degree in Computer Science from the Pune University, Pune, India in 2000 and MCA degree from North Maharashtra University, Jalgaon, India in 2003. She is currently an Assistant Professor at KKWIEER, Nashik. Her research interests include image processing, pattern recognition and computer network.

R. R. Manza received BSc from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India in 1996, M.Sc.(Computer Science) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India in 1998, Ph.D.(Image Data Compression) from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India in 2006.His research interests include Bio Medical Image Processing, Image Compression, Microprocessor, Computer Network, Computer Graphics, Computer Simulation, Computer Vision.