AN AUTOMATIC LEAF BASED PLANT IDENTIFICATION SYSTEM

Nisar Ahmed, Usman Ghani Khan, Shahzad Asif
Department of Computer Science and Engineering, University of Engineering and Technology Lahore, Pakistan.
Corresponding Author: nisarahmedrana@yahoo.com

(Presented at the 5th International. Multidisciplinary Conference, 29-31 Oct., at, ICBS, Lahore)

ABSTRACT: *Plant species identification is an important area of research which is required in number of areas. Design and development of an automatic leaf based plant species identification system is a tough task. The proposed system is based on preprocessing, feature extraction and their weighted normalization and finally classification. We have surveyed contemporary technique and based on their research selected best feature set. Weighted feature normalization is often used in data mining which is applied on this task to improve classification accuracy. Support vector machine is used for classification of plant species by adopting one-vs-all classification approach. The proposed system has provided promising results of 87.40% which will be further enhanced. A completely reliable system for plant species recognition is our ultimate goal.*

Keywords: plant leaf classification, automatic plant species identification, leaf based plant identification, multimedia retrieval

INTRODUCTION

Leaf classification of plants is a useful step for plant health prediction and yield estimation. The recent urbanization and biodiversity loss has made plant identification an important problem for number of professionals such as foresters, environmental protectors, land managers, unprofessional gardeners and agronomists[1]. Plant classification is conventionally done by their floral parts, fruits and leaves by botanist [2]. Selection of flowers and fruits is not a suitable choice of selection for plant identification as they appear for acute interval. Leaves on the other hand are available for longer duration and are available in abundance and happen to be a suitable choice for automatic classification of plants.

Considering the huge amount of species, plant identification is a fairly difficult task even for botanists. Discrimination of different plant species is based on their unique physiognomies. Leaves retain such a discriminative and unique physiognomies thru geometric features and venation architecture. This physiognomic based discrimination has spawned curiosity in the minds of researchers to devise machine based plant identification. In literature different leaf features are used for classification like aspect ratio, circularity, eccentricity, roundness and others. Fourier descriptors[3, 4], wavelets[5, 6], saw tooth pattern, vein structure, color and texture of a leaf [7, 8] and centroid contour distance[9, 10]are also used for leaf identification.

Different varieties of structures are exhibited by the leaf patterns of different plant species. Different approaches has been tried for systematic determination of plant species including leaf shape analysis, more precisely geometric and morphological features. This approach is used in the leaf identification through machines [11].

The proposed approach will automatically identify a plant species by its leaf. The algorithm will extract features based on shape, color and texture of leaf and carefully blend them for optimized results. Feature normalization and dimensionality reduction will be used to counter the effect of dominant feature and increased processing speed. Different classification methods are studied empirically and some best suited classification algorithms will be used for optimized speed and accuracy.

The organization of the rest of this paper is as follows. Section-II provides the details of five different datasets which can be used for evaluation of our proposed algorithm. Section-III provides a literature survey of contemporary research in the area of leaf based plant identification and a comparison table to discuss the features, classifier and other parameters along with the accuracy of the proposed algorithm. Section-IV provides implementation methodology which contains pre-processing, segmentation, feature extractions, feature normalization, dimensionality reduction and classification. We have used careful combination of shape, color and texture based features to obtain precise classification results.

MATERIALS

There are number of dataset available for plant leaf classification task.

Flavia Dataset: This dataset is used by various researchers to demonstrate the plant leaf identification task. It contains highly constrained plant leaf images with truncated stem captured with a white background. This dataset contains plant leaves of 32 different species.

UCI Plant Leaf Dataset: It is a multivariate plant leaf dataset created by Pedro F. B. Silva of Porto University. The dataset contains plant leaves of 40 different plant species and consists of 340 instances. This data is also used by several researchers for plant leaf classification task.

UCI One-Hundred Plant Species Dataset: This dataset is collected by James Cope of Royal Botanic Gardens. It covers one hundred plant species with 16 images of each species comprising of 1600 plant leaf images.

Herbarium Dataset: This data is a high resolution plant leaf images of more than one quarter known plant species. The data set contains 90,000 leaf images. Images used in this dataset has no fixed dimension, aspect ratio or alignment and images are captured with a plain background.

LeafSnap Dataset: This dataset consists of the leaf images taken from two different sources. The lab version consists of over 23000 samples imaged in controlled environment. The field version consists of 7719 samples of images taken in real environment with different devices and ambient conditions. This dataset covers 183 different plant species.

LITERATURE SURVEY

Plants are primary food source and play critical role in almost all food Chain of any ecosystem. This fact made plant taxonomy an open research area for many researchers and this become much easy through automatic plant species identification using leaf descriptor. Authors in [11] have designed a system based on leaf shape descriptor to identify a plant. The method used in their study is application of Zernike moments (ZM) and Histogram of Oriented Gradient (HOG) method as a shape descriptor. This method is experimented on dataset of 50 different plants for accuracy of proposed technique and it resulted in 84.66% accuracy for ZM and 92.67% accuracy for HOG. So overall results obtained from HOG are more satisfactory as compare to ZM as HOG generates more robust shape descriptor features. And using these two feature extraction method with any other better classifier can improve the performance of proposed system.

Plants are mainly classified using their leaf shape. According to others [8], not only the shape of leaf but its color and texture are also important elements which are normally not considered in other existing methods. Authors have proposed a modified system which results in high accuracy of 93.75% which is somehow better than original work with accuracy of 90.312%. The proposed method uses Fourier descriptors, slimness ratio, roundness ratio, and dispersion for shape and for color moments it uses mean, standard deviation, and skewness. It also has used lacunarity for texture and after completing feature extraction, a classifier is applied which is Probabilistic Neural network (PNN) classifier for classification which provides results for plant information. The study is conducted on flavia dataset; 32 different kinds of leaves and can be further extended by adding some other features.

Puja *et al.* [12] designed a system for leaf recognition which is pre-step for plant disease identification as mainly plant diseases are identified through their leaves. The main issue is due to high diversity in plant species which made it complex. To build such a system authors have used to classifiers named as principal component analysis (PCA) and support vector machine (SVM). The method is completed in two phases which are leaf segmentation and classification. The method used 5-binary-classifier to identify plants using SVM and PCA is used to generate feature vector. Furthermore, it used scaling and normalizing of leaves which made it computationally efficient for images of varying size and shape. The author's contribution to classify leaves using both SVM and PCA classifier results in high accuracy of 77.9661%. The accuracy of system also lies in use of Eigen values and SVM for classification process. The study can be further modified by introducing more agricultural classes and also with more features.

Aakif *et al.* [13] have proposed an algorithm for the identification of plants using three separate stages which are pre-processing of dataset then to apply feature extraction which finally contribute to classification of plant. The authors have used features of leaf for identification and classification which are morphological features, Fourier descriptors and shape-defining feature. The obtained values are from these leaf features are used as an input vector for artificial neural network (ANN). The dataset used for this research is based on training through 817 samples of leaves from 14 different fruit trees. The research is experimented for effectiveness on Flavia and ICL datasets and for both these datasets it gives 96% accuracy.

Plant tip and base are important shape parameters and their accurate determination is difficult but can provide significant insight to the plant recognition. Others [10] Centroid Contour Gradient (CCG) is more robust then Centroid Contour Distance (CCD) in tip and base detection of leaf. CCG works by calculating gradient between boundary point pairs of interval angle. The tip extraction accuracy of the proposed algorithm is compared with CCD and a significant improvement from 80.30% to 99.47% is achieved. They have employed Feed-forward Back-propagation algorithm for classification and achieved a classification accuracy of 96.60% as compared to CCD with accuracy of 74.4%.

The authors of [14] have presented a Scale-Invariant Feature Transform (SIFT) based leaf recognition technique. The key descriptors of SIFT are used to detects corners for use in classification. Their proposed scheme is contrasted with two Curvature Scale Space (CSS) based systems. The performance accuracy of SIFT descriptor based approach is 87.5% whereas the accuracies of other two techniques are 71% and 91% respectively. They have used Flavia dataset for leaf image classification.

The research presented by use shape features to contrast and compare the classification performance of three different classifier [15]. They have used five basic geometric features, and 10 morphological features along with leaf vain features. They have done the classification using Probabilistic Neural Networks (PNN) in combination with Principal Component Analysis (PCA) for feature space reduction. The accuracy using this approach was 91%. The second technique is based on Fourier Moments and its accuracy was 71%. The third approach was based on Support Vector Machines (SVM) in addition with Binary Decision Trees (BDT) to deal with multiclass scenario. The efficiency of SVM along with BDT was 96%.

Gulhane *et al*. [16] conducted a research in which they have diagnosed the disease of cotton leaf. For this purposed they have used the two techniques one is Principle Component Analysis and K Neighborhood Classifier. They have used Green channel for identification of Blight disease and shortages of elements can be well reflected by green channel. Aim of this study is to identify the various symptoms of Blight disease which may differ from the normal psychological functioning of plant. Human assistance can be wrong in predication of disease on cotton leaves for this purpose a system is required which can truly identify all cotton leaves diseases. Their proposed technique increases the overall efficiency up to 28%. This technique results in 95% more accuracy which is more than 14% as compare to original manual method.

Pujari *et al.* [17] conducted a research in which they have used image processing technique in order to identify fungal disease symptoms on different agriculture crops. Images are not sufficient in diagnoses of fungal disease which may leads to invalid diagnoses and wastage of time. Early detection of disease can be very useful for this purpose. In order to help farmer their proposed technique is based on the early detection of fungal disease and related symptoms. Many tasks are involved in their proposed methodology. Firstly, they acquits an image than preprocessing of that image is involved further feature selection is intricate than methodology is developed for identification of fungal disease and lastly development of architecture CVS.

Table 1 Comparison Table of Contemporary literature

ID	Paper Title	Methodology	Features	Dataset	Classification	Evaluation	Pros/Cos
1	Identification of the Plants Based on Leaf Shape Descriptors	The study is based on plant identification system based on leaf shape descriptor and for this authors have used HOG and ZM as feature extraction and Euclidean' minimum distance classifier for evaluation of accuracy for different 50 leaves.	Zernike moments (ZM) and Histogram of Oriented Gradient (HOG) method	50 plants datase(VISLeaf database)	'Euclidean' minimum distance classifier	84.66 and 92.67 % accuracy for ZM and HOG, respectively,	**Pros:** High Accuracy and speed. **Cons:** Not good for Multidimensional and sparse data. Small dataset used.
2	Leaf Classification Using Shape, Color, and Texture Features	The research is based on Leaf shape, Vein, color, and texture features are used to classify leaf using PNN classifier after applying feature extraction and experimented on 32 different kinds of leaves.	**Shape features, Fourier descriptors Mean, std. dev. Of and skewness,** lacunarity	Flavia dataset (32 kinds of plant leaves)	Probabilistic Neural network (PNN) classifier	93.75% accuracy, so better as compared to original work that gives 90.312% of accuracy.	**Pros:** High Accuracy and efficiency. Multiple descriptors. **Cons:** Several features. High cost (memory and Computation). Small dataset.
3	Automatic Agricultural Leaves Recognition System	The methodology is based on two phases, first segmentation of leaves and then classification using SVM and eigen values. PCA is used to generate feature vector for classification. The study is experimented on 200 images of four different kinds of leaves. Eigen values Database of 200 images of four different types of leaves principal	Principal Component Analysis, Eigen Values	200 images dataset of four plant species	SVM-PCA	77.9661% accuracy	**Pros:** High Accuracy Fast. Efficient. **Cons:** Small dataset Limited number of classes
4	Automatic classification of plants based on their leaves	The methodology used in this research is to use leaves features for identification of plants which are Fourier descriptors, morphological (FD's) and shape defining-feature (SDF). Then values from these feature are used as an input vector for artificial neural network (ANN) which classify plants. Furthermore it uses dataset of 817 samples from 14 different fruit trees and also applied on Flavia and ICL datasets for	Fourier descriptors	Dataset of 817 samples of leaves from 14 different plants, Flavia dataset, ICL Dataset	Artificial neural network (ANN).	Accuracy 96%	**Pros:** High Accuracy. Flexibility in application. **Cons:** Cannot be used when dataset is updated requires retraining. Results are sometime unclear due to VS dimension.

		effectiveness.					
5	Leaf Recognition using Contour based Edge Detection and SIFT Algorithm	The research is based on SIFT based leaf recognizition. The algorithm is compred with two Curvature Scale Space based algorithms and have found to have greater accuracy from one of them.	SIFT, CSS	Flavia dataset (32 kinds of plant leaves)	No classification method mentioned	87.5% classificatio n accuracy, CSS based techniques shown 71% and 91% accuracy.	**Pros:** Robust to misalignment of leaf. Fewer false-positive and false-negative points. **Cons:** Average efficiency and accuracy.
6	Recognition of Leaf Based on Its Tip and Base using Centroid Contour Gradient	The research is mainly focused on Centroid Contour Gradient to capture the tip and base of the leaf. A comparison has been made with Centroid Contour Gradient and displayed its accuracy. Five plant species are used for leaf recognition task and Feed-forward Back-propagation is used as a classifier.	CCG, CCD	Universit i Teknolo gi Malaysia (25 plant species with a total of 250 samples)	Feed-forward Back-propagation	99.47% accuracy of feature and 96.6% accuracy in classificatio n as compared to 80.30% feature and 74.4% classificatio n accuracy for CCD method.	**Pros:** High Accuracy. Comparison to existing technique. **Cons:** Small dataset. Only one shape feature is used (Focused to CCD only).
7	SVM-BDT PNN and Fourier moment technique for classification of leaf shape	The research is based on the use of three different leaf classification techniques and their comparison. The first method use PNN with PCA, the second method is based on SVM using BDT and the third method is based on Fourier moments. The SVM based on BDT has shown maximum efficiency as well as accuracy.	Basic Shape Features: Diameter, Physiologi cal length, Physiologi cal width and perimeter. Morpholo gical Features: 7 different features Vain features	Flavia dataset (32 kinds of plant leaves)	SVM-BDT, PNN-PCA, Fourier Moments	96% accuracy with SVM-BDT, 91% accuracy with PNN-PCA, 62% accuracy with Fourier Moments	**Pros:** Good Comparison. High Accuracy with the proposed modification to SVM **Cons:** Only based on shape and morphological features.
8	Diagnosis of Diseases on Cotton Leaves Using Principal Component Analysis Classifier	The study is based on identification of disease of cotton by analyzing color of leaves by applying PCA/KNN classifier which results in accurate and efficient results. 110 cotton samples are used for testing using green channel of leaf image.	Eigen Value, Cosine Distance	Dataset of 110 Samples of cotton leaves	Principle Component Analysis and K Neighborhoo d Classifier	Accuracy 95% more than 14% of original manual method, Efficiency increased up to 28%	**Pros:** High Accuracy. High Efficiency. **Cons:** Limited application(Only Cotton) Based on only one feature i.e. color
9	Identification and Classification of Fungal disease Affected on Agriculture/Hort iculture Crops using Image Processing Techniques	Three different methodologies are used for fungal disease identification of three different types of crops. Nearest Neighbor classifier is used for fruit crops, SVM is used for cereal crops, PNN is used for	Euclidean distance, Discrete Wavelet Transform , Local Binary Patterns	dataset is obtained from plant patholog y departme nt of UAS, and UHS INDIA	ANN, PCA KNN, Neuro-KNN, PNN and SVM classifier	85%Accurac y forcereal crops, 94.08% accuracy for fruit crops, 83.17% accuracy for commercial crops and 91.54%	**Pros:** High Accuracy High Effectiveness Pre-Detection **Cons:** Different method for different crop High variability in outdoor condition.

		commercial crops and ANN, PCA KNN, Neuro-KNN are used for vegetable crops.				accuracy vegetable crops	
10	Comparative Study of Leaf Image Recognition with a Novel Learning-based Approach	The research is based on a comparative research study and subsequent implementation of a sparse model based dictionary learning approach for leaf recognition. They have compared their approach with general bag-of-words based approach and shown a superior performance.	Sparse modeling based dictionary learning through different features. Visual-bag-of-words based approach.	Flavia dataset (32 kinds of plant leaves)	Dictionary based and bag-of-words based classification using SVM	Bag of word based approach has shown 94.38% accuracy. Sparse modeling based approach has shown 95.47% accuracy.	Pros: High accuracy. Demonstrated performance by two approaches. Adaptability.. Robust to inaccurate features. Compact feature representation. Cons: Limited number of feature. Less feature diversity.

METHODS

The proposed algorithm will consist of six distinct stage of processing towards the plant species identification. Figure 1 provides a depiction of these steps. The detailed operations of these steps are explained in the following sections.

A. Pre-Processing

Pre-processing is a useful step in plant leaf recognition because it will increase the overall accuracy of the algorithm. There is some inherent noise in every digital noise which may arise due to thermal noise in sensor or improper illumination. A suitable noise reduction algorithm will result better classification accuracy. Moreover, the images are not captured with same cameras or the test images may come from camera of a portable device. We have planned to incorporate color features so the uniformity of color profile is a requirement. RGB is a device dependent color profile and processing in this domain will affect the results based on color features. L*a*b is a color space in which a & b provide device independent color information and L is the luminosity of the image. All the images will be converted to L*a*b color specs before further processing.

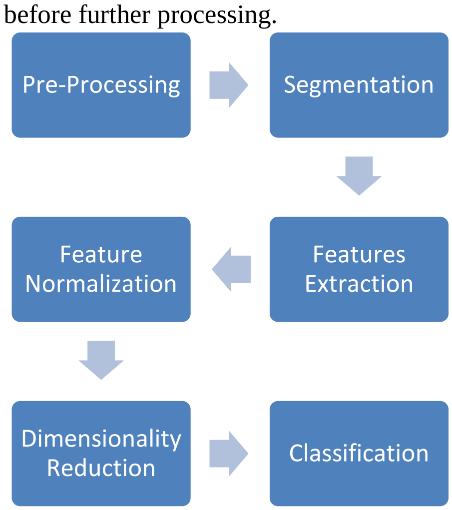

Figure 1 Stages of Plant identification Algorithm

B. Segmentation

Image segmentation is about to separate foreground information from background information. The region of interest (ROI) is extracted for use as a special attribute. Segmentation is a difficult process which consist of number of steps. The segmentation process is defined as identification of spatial area over the image.

To process image for shape features extraction we need to get a binary image with only leaf edges. The image will be converted to grayscale and then binarization is performed using Otsu's method [18]. A rectangular smoothing filter with 3x3 kernel is used to remove noise. To obtain the boundary of the leaf, we will convolve the image with 3x3 kernel of following parameters.

$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

The process of preprocessing and segmentation is pictorially displayed in the figure 2. The final image for shape feature extraction contains the 1-pixel wide boundary of the leaf image.

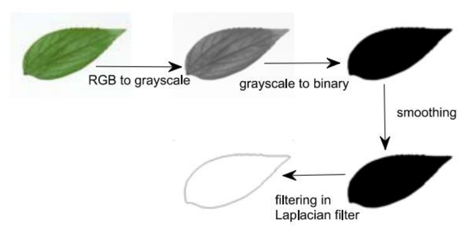

Figure 2 From leaf image to leaf boundary.

C. Feature Extraction

Leaf images can be classified either by color, shape, texture or a suitable combination of any of these[11]. We will exploit suitable features from all three of these by using plain background leaf images. These extracted features will be further processed to be used for classification.

The shape features are mainly related to geometry of leaf and they are also referred to as morphological features. We expect to extract 15 different shape features for further processing and the most suitable features will be retained in the final feature set [19]. The details of these 15 features is provided below.

Diameter: The diameter of the leaf is defined as the longest distance between any two points on the leaf margin.

Physiological Length: The distance between two terminal points of the leaf is defined as physiological length. These terminal points will be indicated by human by clicking on the main vein at the end of leaf stalk and the terminal point of the leaf.

Physiological Width: An infinite number of orthogonal lines can be drawn on the axis of physiological length. The line with the longest length which is normal to the physiological length axis of the leaf will indicate the physiological width of the leaf. Due to discrete pixel structure of the leaf, it is assumed the line is orthogonal even at 90◦ ±0.5◦.

Leaf Area: The leaf area can be easily calculated by counting the number of pixels falling in the smoothed leaf image.

Leaf Perimeter: Leaf perimeter is simply calculated by counting the number of pixels comprising the leaf margin.

Smooth factor: Smooth factor is calculated to measure the effect of noises to the image area. The binary leaf image is processed with 5 × 5 smoothing filter and 3 × 3 smoothing filter, the difference between the areas of both of these processed images is indicated as smooth factor.

Aspect Ratio: It is defined as the ration of physiological length to the physiological width.

Form factor: Form factor describe the difference between a circle of the diameter of leaf and the leaf itself. It is calculated as$\frac{4\pi A}{P}$, where P is the perimeter of the leaf and A indicates the area of the leaf.

Rectangularity: Rectangularity describes the difference between a leaf and a rectangle of the same dimensions. It is calculated as $\frac{L\times W}{A}$, where A is the area of the leaf and L indicates the physiological length of the leaf and W indicates the physiological width of the leaf.

Narrow factor: Narrow factor is a measure of the narrowness of the leaf and is defined as the ratio of the leaf diameter to the physiological length$\frac{D}{L}$.

Perimeter Ratio of Diameter: This factor measures the spreading of the leaf and is defined as$\frac{P}{D}$, whereas P is the perimeter of the leaf and D indicates the diameter of the leaf.

Perimeter Ratio of Physiological Length and **Physiological Width:** This factor also measures the spreading of the leaf by accounting both physiological length and width. It is defined as$\frac{P}{L\times W}$, whereas P indicates the perimeter of the leaf and W indicates the width of the leaf.

Vein features: Vein features are calculated be performing morphological operations on the grayscale image with disk-shaped element of radius 1, 2, 3 and 4. The processed image by morphological operations is subtracted by the margin to obtain the vein structure of the leaf. The areas of the remaining pixels is denoted as $Av_1, Av_2, Av_3\ and\ Av_4$ corresponding to their morphological filter radius. The resultant 5 features are thus obtained by calculating the values of $\frac{Av_1}{A}, \frac{Av_2}{A}, \frac{Av_3}{A}, \frac{Av_4}{A}\ and\ \frac{Av_4}{Av_1}$.

Fourier Descriptors:

Fourier descriptors provides a way to encode an image boundary by mapping every pixel position $(x + y)$into a complex number $(x + iy)$Inverse Fourier transform can be used to recover the original image, however fewer terms can result into a simplified and smooth image. These descriptors can be used for leaf shape encoding using less number of terms which can result into simplified leaf shape for easier identification. Fourier descriptors inherit some properties of Fourier transform such as translation invariance, scaling and rotation.

Fourier descriptor of a leaf boundary can be calculated as:

Record the coordinate values of each pixel sequentially, moving clockwise along the shape.

Construct a complex-valued vector using the recorded coordinate values i.e. $(x + y) \rightarrow (x + iy)$.

Take the DFT of the complex valued vector.

Inverse Fourier transform is used to obtain the original shape of the leaf. The shape can be constructed by using less number of terms (low frequency) which will result in simpler and smooth shape. Many shapes can be approximated with a small number of terms resulting in suitability of Fourier descriptors as an important shape features.

D. Feature Normalization

Normalization of features is a necessary step before dimensionality reduction or classification as the features are extracted by different approaches and their values (ranges) can vary significantly [1]. In the absence of a normalization step, the features with larger values will have a stronger influence during classifier and they will dominate the classification results. Normalization will check maximum and minimum values of each feature and will normalize its value to get a uniform range of values.

E. Dimensionality Reduction

Principal Component Analysis (PCA) is used to reduce the dimension of feature vector by orthogonalization of the feature space [20]. The PCA present the original data as the linear combination of certain irrelevant linear variables. PCA transform the data into a new coordinate systems in such a way that the largest variance by projection of data is placed at first coordinate and the second largest variance is placed at second coordinate and so on.

Each of these new coordinate is referred to as principal component of the original data. First few principal components will be taken out which contribute to almost 95% of feature space to make a new feature vector. This will provide us increased computational speed at the cost of very small decrease in accuracy.

F. Classification

The final stage of the work is classification which define the whole model. This portion describe various classification algorithms which may be used to classify the plant species based on the extracted features. In our scenario, we have employed Support Vector Machines (SVM) which is a popular linear classifier with good accuracy.

G. Support Vector Machines

SVM provide a solution to binary classification problem and has emerged as a powerful technique for learning from data. It is originated from Vapnik's statistical learning theory [21] which formulate the learning as a quadratic optimization problem with global optimum. It maps the input data to a high-dimensional feature space through some nonlinear transformation which is separated by optimal hyperplane. The hyperplane is constructed by separating the negative and positive classes with maximum margin. SVM first transform the input data into a higher-dimensional space by using a kernel function and an optimal hyperplane is constructed between two classes to separate the data in transformed space. The data vectors lying nearest the separating hyperplane are called the support vectors [22]. The SVM estimates a function for classifying data into two classes [21]. Nonlinear support vector classifier can be constructed by replacing the inner product (x, y) with $K(x, y)$. In the following equation, $f(x)$ determines the membership of x.

$$f(x) = sign\left(\sum_{i=1}^{j} a_i\, y_i K(x_i x) + b\right)$$

SVM can be used to construct variety of learning machines based on the selection of different kernels.

H. Multiclass SVM

SVM is inherently a binary classifier which classifies data into two different classes. For leaf based plant identification task, problem of classification involves more than two classes. There are several methods to deal with the problem of multiclass classification problem. The approach used in this study is one-vs-all and choose the class which classify data with maximum margin. In this approach, a single classifier is trained for every class with test cases of that class as positive and all other as negative.

RESULTS

The proposed method is tested on Flavia data set which contains leaf of 32 plant species with 1900 samples. Out of this total sample, 1800 were used as training set and 100 were reserved for testing. All the input leaf images were either 1600×1200 or 800×600 which were resized to the later dimension. The feature set is based on 12 shape features, 5 vain features and Fourier descriptors. A one-vs-all implementation of multiclass SVM is used for classification after dimensionality reduction using principal component analysis.

100 test images belonging to different classes were used to test the accuracy of the proposed method. The algorithm resulted in an aggregate accuracy of 87.40%. This performance is satisfactory at this level, however, the accuracy will be improved further by optimizing the feature vector and classifier. Moreover, other classifier such as probabilistic neural network, convolutional neural network and decision tree based SVM will be checked for the same scheme to obtain optimal accuracy and computational speed.

CONCLUSION

In this study, an automatic plant leaf identification scheme is proposed for facilitation of farmer, novice botanist and hobbyist. This scheme will be further extended to leaf image based plant disease identification. We have performed a literature survey of 10 latest papers in this domain and analyzed the pros and cons of their approaches and based on that survey selected best suitable feature set and classifier. We have also introduced feature normalization and weighting which is beneficial in emphasizing the important features. The proposed scheme has resulted in an aggregate accuracy of 87.40% and will be improved even further by employing the below mentioned approaches.

A. Future Work

Improved segmentation by employing thresholding, region growing and clustering based approaches.

Extension of feature vector by adding better shape features and some texture features.

Employing weighted normalization by testing the retrieval accuracy of individual feature and assigning more weight to it in feature vector.

Testing of other classifier or implementation of binary tree based SVM.

Selection of first five classifier outputs and weighting them accordingly to determine the confidence based class label for reduction of false classification.

REFERENCES

1. Mzoughi, O., et al. *Advanced tree species identification using multiple leaf parts image queries*. in *Image Processing (ICIP), 2013 20th IEEE International Conference on*. 2013. IEEE.
2. Mishra, P.K., et al. *A semi automatic plant identification based on digital leaf and flower images*. in *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*. 2012. IEEE.
3. Neto, J.C., et al., *Plant species identification using Elliptic Fourier leaf shape analysis.* Computers and electronics in agriculture, 2006. **50**(2): p. 121-134.
4. Singh, K., I. Gupta, and S. Gupta, *Svm-bdt pnn and fourier moment technique for classification of leaf shape.* International Journal of Signal Processing, Image Processing and Pattern Recognition, 2010. **3**(4): p. 67-78.
5. Gu, X., J.-X. Du, and X.-F. Wang, *Leaf recognition based on the combination of wavelet transform and gaussian interpolation*, in *Advances In Intelligent Computing*2005, Springer. p. 253-262.
6. Im, C., H. Nishida, and T.L. Kunii. *Recognizing plant species by leaf shapes-a case study of the acer family*. in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*. 1998. IEEE.
7. Rashad, M., B. El-Desouky, and M.S. Khawasik, *Plants images classification based on textural features using combined classifier.* International Journal of Computer Science and Information Technology, 2011. **3**(4): p. 93-100.
8. Kadir, A., et al., *Leaf classification using shape, color, and texture features.* arXiv preprint arXiv:1401.4447, 2013.
9. Wang, Z., Z. Chi, and D. Feng. *Shape based leaf image retrieval*. in *Vision, Image and Signal Processing, IEE Proceedings-*. 2003. IET.
10. Bong, M.F., G. bin Sulong, and M.S.M. Rahim, *Recognition of Leaf Based on Its Tip and Base using Centroid Contour Gradient.* IJCSI International Journal of Computer Science Issues, 2013. **10**(2).
11. Salve, P., et al. *Identification of the Plants Based on Leaf Shape Descriptors*. in *Proceedings of the Second International Conference on Computer and Communication Technologies*. 2016. Springer.
12. Puja, D., M. Saraswat, and K. Arya. *Automatic Agricultural Leaves Recognition System*. in *Proceedings of Seventh International Conference on*

Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). 2013. Springer.

13. Aakif, A. and M.F. Khan, *Automatic classification of plants based on their leaves.* Biosystems Engineering, 2015. **139**: p. 66-75.
14. Lavania, S. and P.S. Matey. *Leaf recognition using contour based edge detection and SIFT algorithm*. in *Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on*. 2014. IEEE.
15. Krishna, S., G. Indra, and G. Sangeeta, *SVM-BDT PNN and Fourier Moment Technique for Classification of Leaf Shape.* International Journal of Signal Processing, Image Processing and Pattern Recognition, 2010. **3**(4).
16. Gulhane, V.A. and M.H. Kolekar. *Diagnosis of diseases on cotton leaves using principal component analysis classifier*. in *India Conference (INDICON), 2014 Annual IEEE*. 2014. IEEE.
17. Pujari, J.D., R. Yakkundimath, and A.S. Byadgi, *Automatic Fungal Disease Detection based on Wavelet Feature Extraction and PCA Analysis in Commercial Crops.* International Journal of Image, Graphics and Signal Processing, 2013. **6**(1): p. 24.
18. Jianzhuang, L., L. Wenqing, and T. Yupeng. *Automatic thresholding of gray-level pictures using two-dimension Otsu method*. in *Circuits and Systems, 1991. Conference Proceedings, China., 1991 International Conference on*. 1991. IEEE.
19. Wu, S.G., et al. *A leaf recognition algorithm for plant classification using probabilistic neural network*. in *Signal Processing and Information Technology, 2007 IEEE International Symposium on*. 2007. IEEE.
20. Shlens, J., *A tutorial on principal component analysis.* arXiv preprint arXiv:1404.1100, 2014.
21. Vapnik, V., *The nature of statistical learning theory*2013: Springer Science & Business Media.
22. Begg, R.K., M. Palaniswami, and B. Owen, *Support vector machines for automated gait classification.* Biomedical Engineering, IEEE Transactions on, 2005. **52**(5): p. 828-838.