International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-4 Issue-4, September 2015

Recognition of Medicinal Leaves using PCA & SVM

Sampath Kumar B, Varsha N, Zaara FK

Abstract*— *In pharmacological science and bhaishaja kalpana, the recognition of plant leaves(leaf) is very important to make medicine. A Recognizing plant leaf has so far been an important and difficult task. In this paper leaf recognition system using the support vector machine as a classifier is proposed. The leaf recognition system consists of image acquisition, preprocessing, feature extraction and classification. The preprocessing involves a typical image processing conversion such as transforming color images to gray scale image then binary , smoothing and to contour image etc. In the feature extraction involves geometrical features are extracted from leaf images such as diameter, length, area, and perimeter etc. Total 10 digital morphological features (DMF) including geometrical features. These 10 features are orthogonalized into five principal variables using principal component analysis (PCA). These features are given as input vector to the support vector machine (SVM) for classifying leaves.

Keywords- Digital Morphological Feature(DMF); leaves(Leeaf) Recognition; Principle component analysis (PCA); Support Vector Machine (SVM)

I. INTRODUCTION

Plants play a main role in the environment. There will be no existence of the earth's ecology without plants. Plant has plenty of use in, medicinal, food, industry, and it is also impordent for environmental protection. Plant is one of the most important forms of life on earth. Plants maintain the balance of oxygen and carbon dioxide of earth's atmosphere . The relations between plants and human beings are very close. In addition plants are important means of livelihood and production of human beings. But in recent years people have seriously destroying the natural environments, so that many plants constantly die and even die out every year. Fortunately, people now have realized that it is a terrible mistake and are beginning to take steps to protect plants. So plant recognition is important and difficult task on the earth. Recently, several species of plants are at the danger of extinction. In order to protect plants and to catalogue various species of flora diversities, a plant database becomes very essential [1].

Revised Manuscript Received on 30 September 2015.
***** Correspondence Author

Sampath Kumar B, Assistant Professor, Department of Electronics and Communication Engineering, Gowdara Mallikarjunappa Institute of Technology, Karur Industrial Area, PB Road, Davanagere, Karnataka, India.

Varsha N, Student, Department of Electronics and Communication Engineering, Proudha Deveraya Institute of Technology, Sha Bhavarlal Babulal Nahar Campus, Hosapete, Karnataka, India.

Zaara FK, Student, Department of Electronics and Communication Engineering, Proudha Deveraya Institute of Technology, Sha Bhavarlal Babulal Nahar Campus, Hosapete, Karnataka, India.

There is huge volume of plant species in the world. All over the world, there are currently about 310000-420000 known plant species,
and many are still unknown yet. In order to handle such volumes of information, development of a rapid and competent classification technique has become an active area of research.[2] Moreover, Along with the conservation of the plants, recognition of plants has also become essential to exploit their medicinal properties and using them as sources of alternative energy sources like bio-fuel.

There are many types to recognize plants. That is based on leaf, root, fruit and stem we can recognize the plants, computer vision and pattern recognition techniques have been applied towards automated process of plant recognition .[3] The morphological features of leaves are employed for plant classification or in the early diagnosis of Certain plant diseases [4]. Plant recognition is an essential and Challenging task. Leaf recognition plays an important role in plant classification and its key issue lies in whether the chosen features are constant and have good capability to discriminate various kinds of leaves. The recognition procedure is very time consuming. Computer aided plant recognition is still very challenging task in computer vision because of improper models and inefficient representation approaches. Plant recognition by computers automatically is a very important task for agriculture, ayurveda medicine (Bhaishaja kalpana), and pharmacological science.

The main aim of leaf recognition is to evaluate the leaf. Digital Morphological Features . This data is very important in identifying the various classes of plants. JiXiang, Huang and Xiao Feng [5] carried out their investigation on recognizing the known plant species by salient features of the leaf.

The extraction of leaf features from a leaf is a key step in the plant recognition process . This feature extraction process creates a new challenge in the field of pattern recognition[6]. The leaf acquisition from living plant to the computer by automatically has not been implemented.

This paper implements a leaf recognition algorithm. The main steps involved in this work are image pre-processing, feature extraction and the classification. Image pre-processing step involves such as image processing steps as transferring colour image to gray level to binary....to contor of leaf (boundary enhancement), feature extraction phase involve 10 digital morphological features extracted including geometrical features, All feature are extracted from digital leaf image. Ten features are orthogonal zed by Principal Components Analysis (PCA)[7] and are given to the classifier. Support Vector Machine (SVM) [8] used as a classifier, for its fast speed and simple structure.

II. PROPOSED SYSTEM

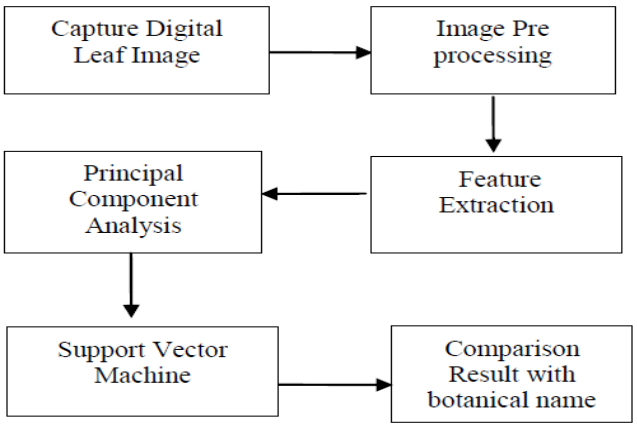


Fig 1 Block diagram of proposed system

The block diagram shows the proposed system for leaf recognition. Which consists of, Image acquisition, Image pre-processing, Feature extraction, PCA and SVM.

2.1 Pre processing

2.1.1 Creation of digital leaf image database

Digital leaf images database can be created by capturing leaf images using digital camera, scanner and also some of leaf images collected from internet. Among these we have used ten species of different plants and each species has ten samples.

2.1.2 Leaf image pre processing

The colors of plant leaves are usually green. However due to changes, the shades and the variety of changes in atmosphere, water and season can cause change of the color of leaf, so the color feature has low reliability. Thus, we decided to recognize various plants by the grey-level image of plant leaf, while ignoring the color information. As a result, only Gray component for each pixel is computed from the color image by that is obtained by using the below formula.

Gray = 0.29989* R + 0.8580* G+ 0:114 *B; (1)

Where R, G, B is the red, green, blue. to the color components of the pixel, respectively. In leaf image pre processing we converted color leaf images into gray level images and then to binary images. After the binary smoothing of leaf images are extracted then the contour of leaves are extracted using morphological operators. Steps of image pre processing is shown in Fig

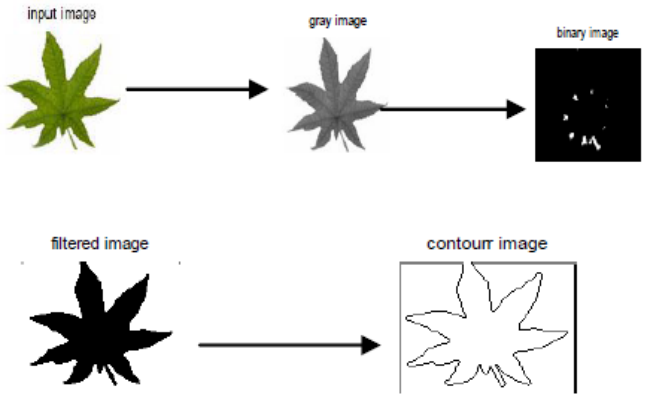


Fig 2 Pre processing

2.2 Feature Extraction

This phase is very important as this phase has greater impact on recognition rate. We need to choose highly discriminate features because these features gives us higher recognition rate This approach uses 10 Digital Mophological features (DMFs) including geometrical features , so that a Computer can extraction feature values automatically.

2.2.1 Digital Morphological Features including geometrical features

*1) Diamete of leaf :*The diameter is the longest distance between any two points on the margin of the leaf. It is represented as D.

2)Length of leaf: The distance between the two terminals of the main vien is the physiological length. It is represented as L.

3) Leaf Area: The value of leaf area is obtained by Counting the number of pixels of binary 1s on smoothed leaf image. Leaf area is denoted as A.

4) Leaf Perimeter: leaf perimeter is obtained by counting the number of pixels present at the leaf margin. It is represented as P.

5) Smooth factor: Smooth factor is the ratio between area of leaf image smoothed by 5×5 rectangular averaging filter and the one smoothed by 3×3 rectangular Averaging filter.

6) circularity: This feature illustrates the difference between a leaf and a circle. It is defined perimete of the leaf margin.

7) Rectangularity : Rectangularity illustrates the similarity between a leaf and a rectangle.

8) Ratio of diameter to length : Narrow factor is the ratio of the diameter D and length L, thus D/L.

9) Rratio of Perimeter to diameter: Ratio of perimeter to diameter, denoting the ratio of leaf perimeter P and leaf diameter D,thus P/D.

10) Rratio of Perimeter to length: This feature is the ratio of leaf perimeter P and length L thus P/L.

In the feature extraction we extract 10 features include DMF and geometrical features.

3.Principal Component Analysis (PCA)

PCA is used to minimize the dimension of input vector of SVM to orthogonality 10 features. PCA is mainly used to give the information of original data as the linear integration of certain linear irrelevant variables.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

It changes the data to new co-ordinate system and each co-ordinate is called principal component ,in which the great variance of data is placed in the first co-ordinate and second greatest variable in the second co-ordinate and so on. To balance the computational complexity we have adopted 5 principal components where in this paper the contribution of these 5 components is 93.6%.To obtain the values of components in new co-ordinate system the mapping. 10 to 5 using the proposed algorithm.

4. Support Vector Machines (SVM)

SVM is a machine essentially used for the purpose of classification,it consists of associated supervised learning techniques. In order to increase the boundary between two data sets SVM generates a two parallel hyper plane. The data point of SVM is denoted as vector P.The SVM is needed to differentiate whether the dimensional hyper planes,if yes then it is classified as a linear classifier. Hence this characteristic feature proves SVM has a potential significance for classification purposes. The whole data in SVM is divided into training and testing where T1 represents training and T2 represents testing. Here the features of T1 are trained using SVM classifier whereas the features of T2 are predicted. Linear classification indicates that the decision surfaces are linear functions of the input vector and therefore are defined using dimensional hyper planes within the dimensional input space. The mathematical representation of the linear classifier surface function of characteristic space is given by,

$$g(x)=\sum nj=I\ aj\ yi\ k(\ x\ ,\ x\ i)\ +\ b \qquad (2)$$

In the above expression (xi,yi) are the two different types of sample collection divided in the sample space, b indicates the classification threshold, k(x,xi) represents the nonlinear kernel function that replace characteristics space which satisfies Mercer conditions. The best linear classification surface function is got by giving the best resolve ai where i=1,2,3….,n for the below equation q(a) .

$$g(x)g(x)=\sum nj=i\ ajyik \qquad (3)$$

$$\sum nj=Iajyi \qquad (4)$$

where i=1,2,……..,n

where Q(a) is convex function. And convex Function is a continuous function whose value at the mid point of each interval in its domain does not exceed the arithmetic mean of its values at the ends of the interval. Because its local optimal solution is global optimal solution, the solution is unique. There are no other possible solutions with better objective function values in the Globally optimal solutions. Hence the best classification function of SVM can be given as:

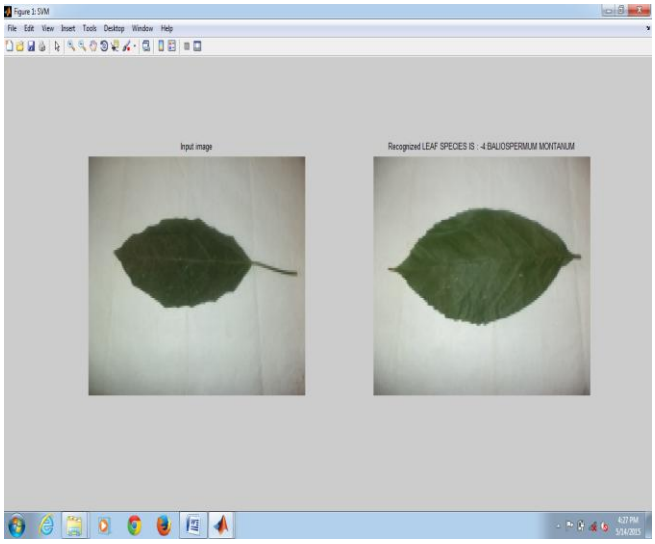
$$f(x)=sgn(g(x))=sgn\{\sum\downarrow(j=1)Tna1jy1jk \qquad (5)$$

$$sgn\{\sum l(j=1)Tna1jy1jk(x,xli)+b\}\} \qquad (6)$$

In order to build a SVM classifier, a kernel function and its parameters is needed. Using a kernel function, SVMs are a substitute training technique for polynomial, radial basis Function and multi-layer perceptron classifiers.

III. EXPERIMENTAL RESULTS

The computer can automatically classify 10 kinds of medicinal leaf images. The performance of the proposed approach is evaluated based on the accuracy and execution time. The accuracy obtained by SVM in medicinal leaf data set is 95 %



3.1 Execution Time

It is observed that in medicinal plants dataset, k-NN approach takes 3.6 seconds for execution where as the proposed approach takes only 0.7 seconds.

Table 1 Executin time

Dataset	Execution time in sec	
	k-NN	SVM Classification
Medicinal leaf	3.6	2.2

Table 2 Leaf botanical names and common names

Botanical Names	Common Names
Ficus Religiosa Linn	Peepal Tree
Ficus Racemosa Linn	Country Fig
Hydnocarpus Anacardium Linn	IndianMarketing nut Tree
BaliospermumMontanum	Red physic nut
Terminalia Arjuna	Arjuna Terminalia, White Murda
Vasanam Pentium	Tulaasi
Celastrus paniculatus	Black Oil Tree

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Mimusops elengi Linn	Bullet WoodTree
Calophyllum inophyllum	Beauty Leaf
Vasanam pentinum	Dasavala

IV. CONCLUSION

A new approach of medicinal leaf recognition, an efficient machine learning approach for leaf recognition is presented in this paper. The approach consists of mainly three steps namely the Image preprocessing, feature extraction and the classification . The computer can automatically classify 10 kinds of medicinal leaf through the digital Cameras or scanners. 10 Digital Morphological Features (DMFs) including geometrical features are extracted In the feature extraction step. And processed by PCA to form the input vector of SVM. SVM classifier is adopted for the classification approach as it has better accuracy, fast training speed and simple structure. The performance of the proposed approach is evaluated based on the accuracy and execution time.

REFERENCES

1. Arun Priya C, Balasaravanan T , and Antony Selvadoss Thanamani "An Efficient Leaf Recognition Algorithm For Plant Using Support Vector Machine", International Conference On Pattern Recognition,Informatics and Medical Engineering(PRIME-2012)
2. Jyotismita Chaki, and Ranjan Parekh, "Plant Leaf Recognition using Shape based Features and Neural Network classifiers," International Journal of Advanced Computer Science and Applications (IJACSA),2011,vol.2,no.10.
3. J. Pan, and Y. He, "Recognition of plants by leaves digital image and neural network," International Conference on Computer Science and Software Engineering, 2008, vol. 4, pp. 906 – 910.
4. N. Kumar, S. Pandey, A. Bhattacharya, and P.S. Ahuja, "Do leaf surface characteristics affect agro bacterium in fection intea," J. Biosci.,vol.29,no. 3, 2004, pp. 309–317.
5. Ji- Xiang Du, De-Shuang Huang, Xiao Feng Wang, and Xiao Gu, "Computer-aided plant species identification (capsi) based on leaf shape matching technique," Transactions of the Institute of Measurement and Control,vol.28,2006,pp.275-284.
6. J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition," Applied Mathematics and Computation, 2007,vol.185
7. J. Shlens, "A tutorial on principal component analysis",2005 [Online]. Available: http://www.cs.cmu.edu/_elaw/papers/pca
8. N. Cristianini, and J. Shawe-Taylor, "An Introduction to Support Vector Machines", Cambridge University Press, 2000.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)