ISSN XXXX XXXX © 2018 IJESC

Research Article Volume 8 Issue No.5

Recognition of Plants by Leaf Image using ANN System

Supriya Sameer Nalawade [1], Shreya Vilas Hule[2], Heenali Vasudev Korgaonkar[3]
Assistant Professor[1], Lecturer[2, 3]
Department of Computer[1], Department of Electrical [2, 3]
Mumbai University, SSPM College of Engineering, Harkul Budruk, Maharashtra, India

Abstract:
Here we are implementing Automatic leaf recognition for plant using PCA and neural network classifier. We are using different features of leaf that are extracted using image processing techniques. These features are then orthogonalised using PCA for reducing dimension of the feature space. Once the features are extracted we can use them to classify the leaves of different plants. For classification we are using artificial neural network.

Keywords: Digital leaf image, feature extraction , PCA , ANN

I. INTRODUCTION

Many researcher have been tried to classify a leaf in the past using k-Nearest Neighbor Classifier (k-NN), Probabilistic Neural Network (PNN), Genetic Algorithm (GA), Support Vector Machine (SVM), and Principal Component Analysis (PCA) . Most of researchers used green color leaves or ignored color information on leaves. Actually, color as features in leaf identification system has been introduced by Man et al. They used the first order, the second order and the third of color moments in HSV color space.

They claimed that the system can recognize 24 categories of plants with the average accuracy up to 92.2%. Several leaf classification systems have incorporated texture features to improve the performance, such as in that used entropy, homogeneity and contraction derived from co-occurrence matrix came from Digital Wavelet Transform (DWT), in that used lacunarity to capture texture of leaf and in that used GLCM.

To implement the classification system, several features were extracted that combined shape, color, vein, and texture features. Then, all those features were processed by using Principal Component Analysis (PCA) to obtained orthogonal features. After that, the results were used in the classification system. The system reads a leaf image as its input, the input leaf image is added to database and will be used for training the neural network.

Selected input leaf image is processed. It will be matched with the images of the leaves in the database. The relevant information about the leaf from the database will be displayed. Some existing system are present, such as cell and molecule biology methods, classification based on leaf image is a better choice.

Take image and convert it into digital signals are low cost and convenient. One can easily transfer the leaf image to a computer and a computer can extract features automatically in image processing techniques. Guyer et al. implemented an algorithm to extract plant/leaf shape features using information gathered from critical points along object borders, such as the location of angles along the border (and/or) local maxima and minima from the plant leaf centroid[11]. Franz et al. identified plants based on individual leaf shape described by curvature of the leaf boundary at two growth stages.[12]

II. PROPOSED SYSTEM ARCHITECTURE

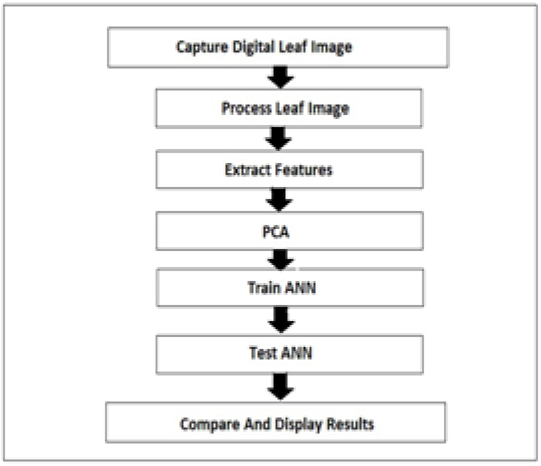

Figure.1. Proposed Methodology for recognition of plant by leaf image using ANN

The leaf recognition system will consist of many steps. The starting will be with pre-processing of the images of the leaves. The pre-processing will smooth the image and make it more suitable for further processing. Next we will extract features from the image. Next we will use PCA for orthogonalise the feature set.

This simply means we are reducing number of features that are used for recognition. This will also decrease the time as well as reduces the complexity of the system. These features are then stored as a database in the system. The database then use to train the artificial neural network by taking training sample and comparing them with dataset. Once the classifiers are trained we can then use it on testing sample. After classification we can get the name of category to which the leaf belongs. Boundary Enhancement When mentioning the leaf shape, the first thing appears in your mind might be the margin of a leaf. Convolving the image with a Laplacian filter of following 3 * 3 spatial mask: we can have the margin of the leaf image. To make boundary as a black curve on white background, the 0 and 1value of pixels is swapped.

Feature Extraction

5 basic features are extracted from 12 morphological features, so that a computer can obtain feature values quickly and automatically (only one exception).

Basics Geometric Features

- Diameter: The longest distance between any two points on the margin of the leaf. It is denoted as D.
- Physiological Length:The distance between the two terminals is defined as the physiological length. It is denoted as Lp.
- Physiological Width: The longest distance between points of those intersection pairs is defined at the physiological width. It is denoted as Wp. Since the coordinates of pixels are discrete.
- Leaf Area: It can evaluate by counting the number of pixels of binary value 1 on smoothed leaf image. It is denoted as A.
- Leaf Perimeter: It means counting the number of pixels consisting leaf margin. Denoted as P

Twelve Digital Morphological Features

In this section we define 12 digital morphological features used for leaf recognition.

- Smooth factor: We use the effect of noises to the area of image to describe the smoothness of leaf image. It is the ratio between the area of leaf image smoothed by 5* 5 rectangular averaging filter and the one smoothed by 2*2 rectangular averaging filter.
- Aspect ratio: Lp/Wp.

where Lp is physiological length and Wp is physiological width

- Form factor: This feature is used to describe the difference between a leaf and a circle. It is defined as (4*\ pi*A)/P2, where A is the leaf area and P is the perimeter of the leaf margin.
- Rectangularity: Rectangularity describes the similarity between a leaf and a rectangle. It is defined as LpWp/A, where Lp is the physiological length, Wp is the physiological width and A is the area of leaf.
- Narrow factor: D/Lp.

where D is diameter and Lp is physiological length

Perimeter ratio of diameter: It the ratio of leaf perimeter P and leaf diameter D, is calculated by P/D.

- Perimeter ratio of physiological length and physiological width: This feature is defined as the ratio of leaf perimeter P and the sum of physiological length Lp and physiological width Wp, thus P/(Lp +Wp).
- Vein features: We perform morphological opening [6] on grayscale image, disk-shaped structuring element of radius 1,2,3,4 and substract remained image by the margin.The results look like the vein. Areas of left pixels are denoted as Av1, Av2, Av3 and Av4 are five vein features. Then we obtain the last 5 features: Av1/A, Av2/A, Av3/A, Av4/A, Av4/Av1.Now we have finished the step of feature acquisition and go on to the data analysis phase.

III. PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) is a statistical method that the main goal is to reduce dimension of data. A N x N pixel image of a leaf, represented as a vector occupies a single point in N2-dimensional image space. Leaf images will not be describe into high image space. Therefore, they can be described by a low dimensional subspace.

Main idea of PCA:

1.To find vectors that best account for variation of leaf images in entire image space.

2.These vectors are called eigenvectors.

3.Construct a leaf space and project the images into this leaf space (eigen faces).

Algorithm

Step 1: Get some data

Step 2: Subtract the mean

In this step subtract the mean from each of the data elements. The subtracted value is the average across each data element. So, all the x values have $\bar{x}$ (the mean of the x values of all the data points) subtracted, and all the y values have $\bar{y}$ subtracted from them.

Step 3: Calculate the covariance matrix

From this new image space of M ϕi images (Each with dimension N x 1), the matrix A is formed with dimension N x M by taking each of image vectors ϕi and placing them in each column of matrix A.[9][10]

A = [ϕ1ϕ2ϕM]

Using matrix A, it is important to set up the Covariance matrix C. This can be given by product of matrix A with matrix A^T The dimension of such covariance matrix will be N x N .[9][10]

$C = A A^T$

Step 4: Calculate the eigenvectors and eigenvaluses of the covariance matrix

As the dimension of this matrix is N x N, which means it will result in N eigenvalues and N eigenvectors. Since the value of N is very large, say 65536 as in above example, it would be better to reduce this overhead by considering matrix $L = AA^T$. The dimension of this matrix will be M x M.

$L = A A^T$

The N eigenvalues obtained from C are same as M eigenvalues with remaining N - M eigenvalues equals zero. Also if x is eigenvector obtained from C then the eigenvectors of L are given by [9][10]

$y = A^T x$

We can make use of this relationship to obtain eigenvalues and eigenvectors of AA^T by calculating eigenvalues and eigenvectors for AA^T . The eigenvectors for C (Matrix U) are obtained from eigenvectors of L (Matrix V) as given below:

U = AV

The matrix V, with dimension (M x M), is constituted by the M eigenvectors of L and matrix U, with dimension (N x M), is constituted by all the eigenvectors of C, and the matrix A is the image space, with dimension (N x M).

Step 5: Choosing components and forming a feature vector

From the covariance matrix obtain the eigenvector and arrange them in order to represent it as eigenvalue, highest to lowest. This gives you the components in order of significance. Now, you can decide to ignore the components of lesser significance. In this case you lose some information, but if the eigenvalues are small, you don't lose much. Therefore final data set will have less dimensions than the original. To be precise, if you originally have η dimensions in your data, and so you calculate η eigenvectors and eigenvalues, and then you choose only the first eigenvectors, then the final data set has only dimensions matrix. Now you need to form a feature

vector, which is just a fancy name for a matrix of vectors. Feature vector is constructed by taking the eigenvectors from the list of eigenvectors, and forming a matrix with these eigenvectors in the columns.[9][10]
|Feature Vector ={eig1,eig2,eig3.....eign}

Step 6: Deriving the new data set
This is the final step in PCA, and also the easiest. Once we have chosen the eigenvectors and formed a feature vector, we simply take the transpose of the vector and multiply it on the left of the original data set, transposed.

Final Data=Row Feature Vector x Row Data Adjust

Image Representation:
Training set of images of size N*N are represented by vectors of size N2

$\Gamma_1,\Gamma_2,\Gamma_3,\ldots,\Gamma_M$

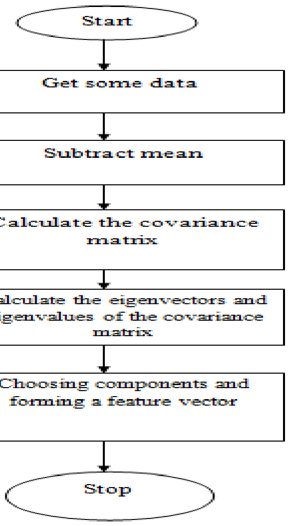

Figure.2. Flowchart of PCA algorithm

IV. INTRODUCTION TO ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a magical • black box trained to achieve expected intelligent process, against the input and output information stream. Thus, there is no need for a specified algorithm on how to identify different plants. Neural networks are used in the automatic detection of leaves diseases. Neural network is chosen as a classification tool due to its well known technique as a successful classifier for many real applications. The training and testing processes are important steps in developing an accurate process model using NNs. The dataset for training and validation processes consists of two parts; the training feature set which are used to train the NN model; whilst a testing features sets are used to verify the accuracy of the trained NN model. Before the data can be fed to the ANN model, the proper network design must be set up, including type of the network and method of training. This was followed by the optimal parameter selection phase. However, this phase was carried out simultaneously with the network training phase, in which the network was trained using the feed-forward back propagation network Training leaf image enter the neural network and according to their class, a back propagation error, spread on the network and correct the weights toward the right values. The input leaf image will classified to the class which has the greatest similarity to it.

V. APPLICATIONS AND FUTURE SCOPE

If plants are at the risk of extinction. So it is very necessary to set up a database for plant protection.This system can be useful for:
Pharmaceutical Industries.
Botanists.
Medical Research Centers.
Farmers and any normal users.
A leaf identification system that incorporate PCA has been developed. Our plant database is under construction. The number of plants that can be classified will be increased. In the further studies we can also identify diseases in plants.
The experiment show that PCA can improve the accuracy of the system. However, some other works will be explored to obtain better performance.

VI. EXPERIMENTAL RESULTS

Fig 3 shows the snapshot of database creation. In the proposed system leaf image can be easily added in the database using MATLB. The training feature sets are used to train the NN model.

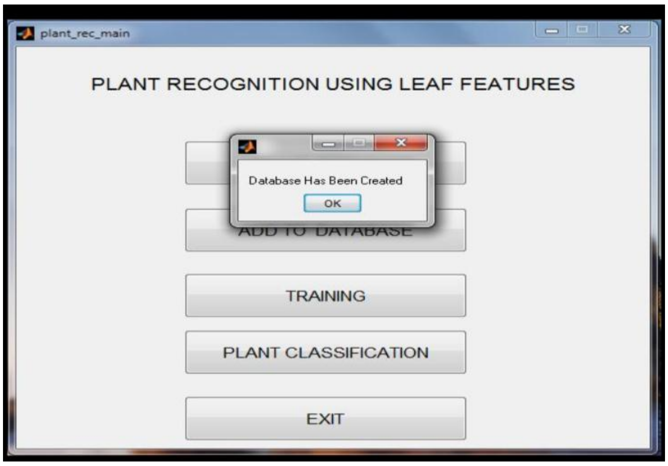

Figure. 3. Creat Database

Fig.4 shows indicates snapshot of the leaf image which is to be recognized is read from given database. Here input is given to the system

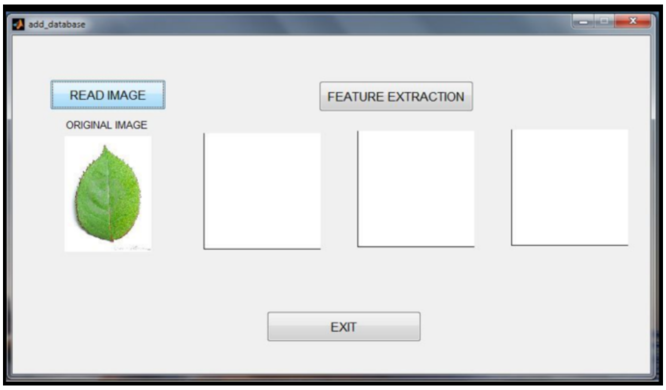

Figure.4. Read image from database

After that leaf image is preprocessed and converted into gray image . From the original image features are extracted using PCA Algorithm as shown in Fig.5

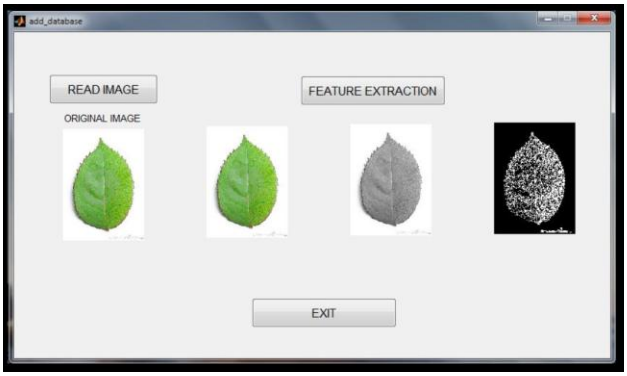

Figure. 5. Feature extraction

Finally proposed system recognized the given input using ANN classifier. Fig.6 represent output window, indicated as given image is a rose leaf.

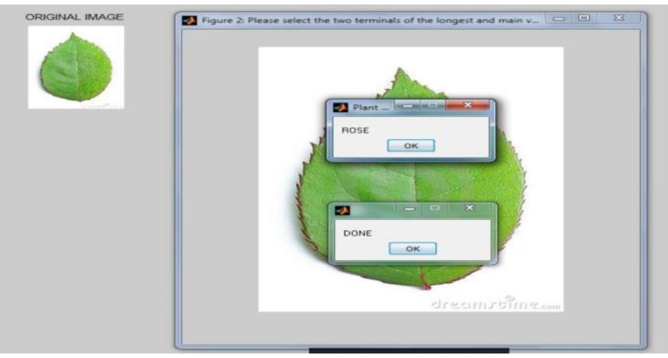

Figure. 6.Output window

In the test case we test that the image other than leaf are not classified.Our system will extract the features of image but do not classify it.This is illustrated in fig. 7

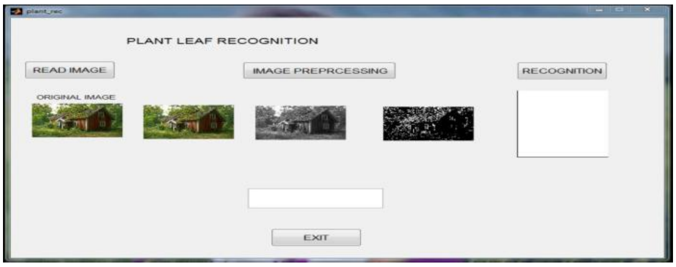

Figure.7. Test case1

In this test case we test that our system only extracts the features of black and white image of leaf. But does not classify the leaf. This is illustrated in fig 8.

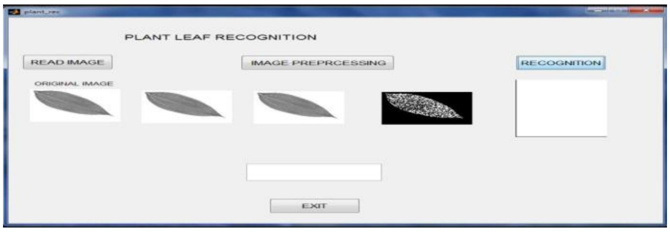

Figure.8. Test case2

VII. CONCLUSION

By the end of this project we were able to design a simple leaf recognition System using matlab Programming. The work described in this research has been concerned with the two challenging phases in image analysis applications which are feature extraction and classification phase. The computer can automatically classify plants via the leaf images. ANN is adopted for its fast training speed and simple structure. Twelve features are extracted and later processed by PCA to form the input vector of ANN. This algorithm is fast in execution, efficient in recognition and easy in implementation. The experiments show that PCA can improve the accuracy of the system.

VIII. REFERENCES

[1].J.-X. Du, X.-F. Wang, and G.-J. Zhang, "Leaf shape based plant species recognition", Applied Mathematics and Computation, vol. 185, 2007

[2].Y. Ye, C. Chen, C.-T. Li, H. Fu, "A computerized plantspecies recognition system", in Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong,

[3].M. T. Hagan, H. B. Demut, and M. H. Beale, Neural Network Design, 2002

[4].D. F. Specht, "Probabilistic neural networks for classification mapping, or associative memory", in Proceedings of IEEE International Conference on Neural Networks, vol. 1, 1988

[5].T. Master, Practical Neural Network Recipes. New York: John Wiley,1993.

[6].R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "Digital Image ProcessingUsing MATLAB". Prentice Hall, 2004.

[7].J. Du, D. Huang, X. Wang, and X. Gu, "Shape recognition based on radial basis probabilistic neural network and application to plant species identification", in Proceedings of 2005 International Symposium of Neural Networks, ser. LNCS 3497. Springer, 2005.

[8]. I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities", Nature, vol. 447, May 2007.

[9].KIRBY, M. AND SIROVICH, L. 1990. Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans. Patt. Anal. Mach. Intell. 12.

[10]. A tutorial on Principal Components Analysis Lindsay I Smith February 26, 2002.

[11]. Guyer, D.E., G.E. Miles, D.L. Gaultney, And M.M. Schreiber, Application Of Machine Vision To Shape Analysis In Leaf And Plant Identification"". Transactions Of ASAE, 1993, Volume 36(1), pp-163:171.

[12]. Franz, E., M.R. Gebhardt, And K.B. Unklesbay, The Use Of Local Spectral Properties Of Leaves As An Aid For Identifying Weed Seedlings In Digital Images. Transactions of ASAE, 1991. Volume 34(20), pp- 682:687.