Orthogonalise Digital Morphological Features Using Principal Component Analysis

Vandana R. Patil[1] , Ramesh R. Manza[2]

[1]Assistant Professor, Department of MCA Engineering, KKWIEER, Maharashtra, India

[2]Associate Professor, Department of CS & IT, Dr. BAMU, Maharashtra, India

---***---

Abstract - ***Plants play an important role in preserving earth ecology and maintaining healthy atmosphere. Most of the food sources are from plants. So it is necessary to establish a system to extract the features of plant. As there are large variety of plant species available in the universe . There are various parts of plant which can be chosen for feature extraction. In this research we choose shape of leaf to extract the features. Every plant leaf has variety of features to extract. This research reduces large number of digital morphological features to five principal components. This research describes procedure to transform raw image to a preprocess image for feature extraction. Further five geometrical features are extracted from preprocessed image and from geometrical features twelve digital morphological features are extracted. Due to large number of digital morphological features it is difficult to proceed for further for plant identification. These twelve digital morphological features are orthogonalise to five principal components using principal component analysis.***

Key Words : Morphological feature, plant leaf classification, feature extraction, digital morphological features, principal component analysis.

1.INTRODUCTION

Plants play an important role in our ecosystem. They are useful for human beings and animals too. Plants provide food, medicine, oxygen and many important substances for living and non living thing. Plants having very broad range of applications in agriculture and medicine. So it is necessary to know which plants are useful or harmful to human being to save the life of human being. Due to the effect of global warming large number of rare plant species are at the boundary of extinction. But it is important and difficult task to recognize the plant species. Plants recognition is useful in agriculture and medicine as well a biological diversity research. It is also useful in tea , cotton and other industries. So it is necessary to develop a database by information technology as soon as possible. To recognize the plant species researcher needs to extract the various features of plant species. To proceed with this, features can be extracted from various parameters of plant like leaf, fruit, flower or stem . In this research we choose shape of leaf parameter to extract the feature. The aim of this research is to transform the digital morphological features of leaf to principal components. As the number of features extracted from leaf are in large number , so it is difficult for classifier to proceed with this to classify the plant species from plant kingdom. The finding of this research is useful for plant classification purpose.

2. PREVIOUS WORK

Global feature and local descriptor are two categories for features of leaf, as stated by Shabanzade et al(2011)[1]. According to C. L. Lee , S. Y. Chen [2] past research in recognizing objects can be broadly classified into two categories : a) contour based and b) region based approaches. The disadvantage of the contour based feature is the difficulty on finding the correct curvature points. Based on the contour of leaf, features were extracted to differentiate species. However contour of leaves have variation even in the same species. For plants identification purpose Wu, et al[3] used shape slimness, defined as ratio of length to width of leaves, shapes roundness, defined as ratio of area of leaf image and perimeter of leaf contour, and shape solidity, defined as ratio of the internal area connecting to valley points and the external area connecting the top points. A paper by Ji- Xiang Du, Xiao-Feng Wang, Guo-Jun Zhang [4] introduce how to extract digital morphological features. The features are extracted from the contours of leaf. The digital morphological features(DMF) generally include geometrical features(GF) and invariable moment features(MF). A paper by Cholhang Im, Hirobumi Nishia and Tosiyasu L. Kunii[5], a method that normalizes shapes of leaves is presented using symmetry of each leaflet with respect to its vein. According to Najjar and Zagrouba [6] had used region based feature for proposed method in order to classifying the leaf. According to C.S Sumathi and A.V.Senthil Kumar in plant classification, the leaf shape plays a significant role. In machine intelligence, the most significant part essential for both decision making and data processing is shape recognition. Valliammal and Geethalakshmi [7], who stated in their publications that leaf image could be categorized based on color, texture, shape or combination of these properties. Later Zhang and Zhang[8] was enhanced that the properties for these features such as surface area, surface perimeter and the disfigurement are inherited from the shape features, variance of red, green and blue channels are belonging to the color features and texture energy. Other research was also used aspect ratio and other basic geometric features to recognize plants. Du, Wang and Zhang used Aspect ratio, rectangularity, area ratio of convex hull, perimeter ratio of convex hull and sphericity [3].

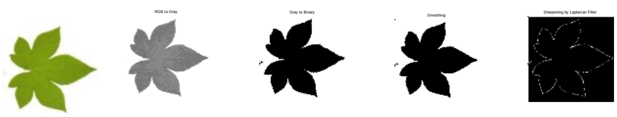
[9] this paper proposes a method for an automatic identification of plant species from low quality pictures of their leaves created using mobile devices. To avoid segmentation of the images local features are used which are scale and rotation invariant.[10] this paper proposed a method that combines polar fourier transform, color moments and vein features to retrieve leaf images based on leaf image. Du et al [11] captured the leaf shape polygonial approximation and algorithm called MDP(modified dynamic programming) for shape matching.

3. METHODOLOGY

Firstly we have converted the 2 dimensional image of plant leaf to grey level image . For each conversion of a color image to grayscale is not unique, different weighting of the color channels effectively represent the effect of shooting black-and-white film with different strategy is to match the luminance of the grayscale image to the luminance of the color image. To convert plant leaf color image to a grayscale representation of its luminance, Firstly we have obtained the values of its red, green and blue(RGB)primaries. Finally only gray component for each is computed from the color image by

Gray = 0.2999 * R + 0.578 *G+ 0.114 * B

Where R,G,B correspond to the red, green and blue color of the pixel respectively. This image is then transformed into a binary image. Further , smoothing filter is applied to the binary image to reduce the noise . The boundaries are gained by applying a laplacian filter. The steps involved in pre-processing technique are illustrated in the following figure



OriginalImage(RGB)→Gray→Binary→Smoothing-→Contours of Leaf

The next step is the digital morphological feature extraction . The aim behind this step is to remove redundancy from the image and to represent by a set of numerical features. These features are extracted from the contours of leaf. The digital morphological features contain basic geometric features and morphological features. Our system extract 5 basic geometric features i.e longest diameter, physiological width, leaf area and leaf perimeter. Our system also extract 12 digital morphological feature are as follows smooth factor, aspect ratio, form factor, rectangularity, narrow factor, perimeter ratio of diameter, perimeter ratio of physiological length and physiological width and 5 vein features : f1,f2,.....f5.

3.1.Five basic Geometric Features

1. Longest Diameter : It is longest distance between any two points on the contours of leaf. It is denoted as D.
2. Physiological length : Here one needs to select two terminal of the main vein of leaf via mouse click. The distance between these two terminal is defined as the physiological length. It is denoted as L_P
3. Physiological Width : Once physiological length is find out i.e line drawn between two selected terminal on the main vein, one can plot infinite lines orthogonal to that line. We consider two lines are orthogonal if their degree is 90^0. Physiological width is consider as longest distance between points of those intersection pairs. It is denoted as W_P
4. Leaf Area : Smoothed leaf image is consider to find out leaf area. Number of pixels having binary value 1 is termed as leaf area. It is denoted as A
5. Leaf Perimeter : Leaf Perimeter is calculated by counting the number of pixels consisting leaf margin. It is denoted as P

B.*Digital Morphological Features*

Based on above mentioned five basic geometric features, we defined 12 digital morphological features

1. Smooth factor : It is defined as the ratio between area of leaf image smoothed by 5 X 5 rectangular averaging filter and area of leaf image smoothed by 2 X 2 rectangular averaging filter.

2. Rectangularity : It is defined as the ratio between physiological length, physiological width and leaf area.
   Thus, $L_P W_P/A$.
3. Aspect ratio : It is defined as the ratio between physiological length and physiological width.
   Thus, L_P / W_P.
4. Perimeter ratio of diameter : It is defined as the ratio between perimeter and diameter.
   Thus, P/D.
5. Form factor : It is defined as 4 Л A/P· where A is area of leaf and P is perimeter of the leaf margin.
6. Narrow factor : It is defined as the ratio between diameter and physiological length.
   Thus, D / L_P
7. Perimeter ratio of physiological length and physiological width : It is defined as the ratio between perimeter and sum of physiological length and physiological width. Thus, $P / (L_P + W_P)$
8. Vein features : We perform morphological opening on grayscale image with flat, disk shape structuring element of radius 1,2,3,4 and subtract the remain image by the margin. Then gray threshold value is computed and with that threshold image is converted to binary. 8/4 connected components of leaf are find out and measured set of properties of each connected component. The result is look like the vein. Areas of left pixel are denoted as A_{V1}, A_{V2}, A_{V3}, A_{V4} respectively. Then obtain the last five features as A_{V1}/A, A_{V2}/A, A_{V3}/A, A_{V4}/A, A_{V1}/ A_{V1}.

3.2. Principal Component Analysis

Principal component analysis was invented in 1901 by Karl Pearson. Principal component analysis is a technique for identifying uncorrelated variable called "principal component". The objective of PCA is to perform dimensionality reduction while preserving as much of the randomness in the high-dimensional space as possible. This method primarily focus on variable reduction procedure. It mainly used when we obtained data on large number of variables and we know that there is some amount of redundancy in obtained data. So it is difficult to proceed with this large number of variables. All the principal components are orthogonal to each other, so there is no redundant information. The principal components as a whole form an orthogonal basis for the space of the data. Mathematically , PCA transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate, the second greatest variance on the second coordinate and so on[14]. Each coordinate is called a principal component. In this research twelve digital morphological features of leaf are reduced to five principal components using principal component analysis. It is observed that variability of data is reduced to a relatively small number of principal components. In result PCA can achieve high dimensionality reduction with usually lower noise that the original patterns.

4. EXPERIMENTAL RESULT

We use ICLleaves dataset . This dataset contains 32 kind of different plant species. Each species includes 20 sample images. Hence there are 400 images with the database.

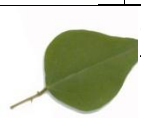
Leaf Image	Diameter	Physiological Length	Physiological Width	Leaf Area	Leaf Perimeter
	345.1623	268.7452	154.0195	3.1099 e+04	799
	542.0793	364.9288	195.4329	5.0621 e+04	1.2531 e+03
	374.6692	302.4169	232.6070	5.2468 e+04	987.3750
	324.8400	224.4037	180.9420	2.8043 e+04	789.8750
	313.6367	266.2724	164.7786	3.5723 e+04	757.5000

Table 1.1 : Results of Geometrical Feature Extraction

Leaf\Feature	Narrow Factor	Perimeter ratio of Diameter	Perimeter ratio of Lp and Wp
	2.2410	2.3149	1.8899
	2.7737	2.3117	2.2363
	1.6107	2.6353	1.8455
	1.7953	2.4316	1.9486
	1.9034	2.4152	1.7573

Leaf Image \Feature	Smooth Factor	Aspect Ratio	Form Factor	Rectangularity
	0.9878	1.7449	0.6122	1.3310
	0.9881	1.8673	0.4051	1.4089
	0.9935	1.3001	0.6763	1.3407
	0.9919	1.2402	0.5648	1.4479
	0.9868	1.6159	0.7823	1.2282

Table 1.2 : Results of Digital Morphological Features

Leaf Image\Principal Component	P1	P2	P3	P4	P5
	0.2396	-0.3099	-2.0805	-1.6455	0.1686
	0.5986	-0.3440	-1.5846	0.2776	0.5432
	1.1065	-0.5549	-2.6279	-0.2858	0.5647
	0.9256	-0.4583	-2.5108	-0.2694	0.5408
	0.9901	-0.2494	-2.4770	0.5253	0.5822

Table 1.3 : Results of Digital Morphological Features

Leaf Image/Feature	F1	F2	F3	F4	F5
	0.0784	0.0223	0.0222	0.0260	0.3317
	0.1175	0.2448	0.3361	0.2819	2.4001
	0.1799	0.2790	0.2916	0.3495	1.9432
	0.1432	0.1650	0.2349	0.2776	1.9387
	0.1107	0.1718	0.2521	0.3101	2.8019

Table 1.4 Results of Vein Features

Table 1.5 : Results of Principal Component Analysis

5.CONCLUSION

In this paper an approach for transforming digital morphological features to principal component analysis is presented. First the images are transformed into grayscale and normalized, then converting grayscale image to binary and binary to smoothing. Using Laplacian filter boundaries are enhanced from binary smooth image. Further 5 basic geometrical features and 12 digital morphological features are extracted from the contours of leaf. Due to the large number of extracted features , it is difficult to classify . So these features are reduced to five principal components. They are further useful for plant identification.

ACKNOWLEGEMENT

We would like to thank the "Intelligent Computing Science Lab., Chinese Academy of Sciences, China" for their support and for providing the images.

REFERENCES

[1] Shabanzade, M.M.Zahedi and S.A.A ghvami, 2011. Combination of local descriptors and global features for leaf recognition signal and image processing. International Journal, vol. 2 : 23-21. DOI : 10 .5121/SIPIJ

[2]C.L.L EE, S.Y. Chen,(2003). "Classification for Leaf Image" , 16th IPPR Conference on Computer Vision , downloaded at www.csie.mcu.edu.tw/`yklee/CVGIP03/CD/Paper/PR/PR-19.pdf.

[3]Q.W.Wu, C.Zhou,C.Wang,(2006) " Feature Extraction and Automatic Recognition of Plant Leaf using Artificial Neural Network ", Advances in Ciencias de la Computation , pp. 5-12.

[4]Ji-Xiang Du, Xiao-Feng Wang,Guo-Jun Zhang. "Leaf Shape based plant species recognition", Applied Mathematics abd Computation 185 (2007) 883-893.

[5] Cholhong Im, Hirobumi Nishida, Tosiyasu L. Kunii . " Recognizing Plant Species by Normalized Leaf Shapes". Vision Interface '99, Trois-Rivieres,Canada, 19-21 May.

[6] Najjar, A. and E.Zagrouba, 2012. Flower image segmentation based on color analysis and a supervised evaluation. Proceddings of the International Conference on Communications and Information Technology (ICCIT) , Jun. 26-28, IEEE Xplore Digital Library, pp : 397-401. DOI : 10.1109/ICCITechnol.2012.6285834

[7] N. Valliammal and S.N. Geethalakshmi, (2011). Automatic Recognitin System Using Preferential Image Segmentation for Leaf and Flower Images, Computer Science and ENGINEERING : An International Journal (CSEIJ),Vol. , No. 4, pp.13-25.

[8] Stefan Fiel and Robert Sablantnig ."Leaf classification using local features."

[9]Abdul Kadir,Lukito Edi Nugroho, Adhi Susanto and Paulus Insap Santosa , " Foliage plant retrieval using polar fourier transform, color moments and vein features", An international Journal(SIPIJ) Vol. 2, No.3, September 2011.

[10]Du. J.X ., Huang, D.S.Wang,X.F., & Gu, X.(2006).: "Computer –aided Plant Species Identification (CAPSI) Based on Leaf Shape Matching Technique". Transactions of the Institute of Measurement and Control, 28(3),pp.275-284

[11] Jing, L.Z.Shawan and L.Jiandu , 2009. "A method of Plant Leaf Recognition based on Locally Linear Embedding and Moving Center Hypersphere Classifier". Procedding of Intelligent Computing Technology and Application Springer-verlag (ASV'09), Uslan, South Korea,PP : 645-651. DOI : 10.1007/978-3-642-04020-769

[12] R.C.Gonzalez, R.E.Woods and S.L.Eddins, Digital Image Processing Using MATLAB. Prentice Hall , 2004.

[13] 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

[14] Krishna Singh, Indra Gupta, Sangeeta Gupta, SVM-BDT PNN and Fourier Moment Technique for classification of Leaf, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 3, No. 4, December, 2010