International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014
ISSN 2091-2730

DETECTION AND CLASSIFICATION OF PLANT LEAF DISEASES

Kshitij Fulsoundar[1], Tushar Kadlag[2], Sanman Bhadale[3], Pratik Bharvirkar[4]

Prof S.P.Godse[5]

[1,2,3,4]Student,[5]Guide Department of Computer Engineering, Sinhgad Academy of Engineering, Pune, Maharashtra, India

Abstract- Medicinal plants used very much in herbalism to study the medicinal properties of the plants. The applications of Near-Infrared Spectroscopy (NIRS) have expanded widely in the field of agriculture, plants and various other fields, but the usage for identification of plant variety is still rare.In this project we describe the development of an Android application that gives users the ability to identify plant species based on photographs of the plant's leaves taken with a mobile phone. At the heart of this application is an algorithm that acquires morphological features of the leaves, computes well documented metrics such as the angle code histogram (ACH), then classifies the species based on a novel combination of the computed metrics. The algorithm is first trained against several samples of known plant species and then used to classify unknown query species. Aided by features designed into the application such as touch screen image rotation and contour preview, the algorithm is very successful in properly classifying species contained in the training library.

KeyWords:Android; Java;Oracle Database;Image Processing

1. INTRODUCTION

Plants play an important role in the cycle of nature. The number of plant species is estimated to be around 400,000 however there still exist many species which are yet unclassified or unknown. Therefore, plant identification is a very important and challenging task. With the rapid progress of information technologies, many works have been dedicated to applying the technologies of pattern recognition and image processing to plant identification. Since leaves are the organ of plants and their shapes vary between different species, the leaf shape provides valuable information for plant identification. The paper aims to describe, leaf identification based on shape. Leaf shape description is the key problem in leaf identification. Up to now, many shape features have been extracted to describe the leaf shape.

In this project, we describe the development of an Android application that gives users the ability to identify plant species based on photographs of the plant's leaves taken with a mobile phone. At the heart of this application is an algorithm that acquires morphological features of the leaves, computers well documented metrics such as the angle code histogram (ACH), then classifies the species based on a novel combination of the computed metrics. The algorithm is first trained against several samples of known plant species and then used to classify unknown query species. Aided by features designed into the application such as touch screen image rotation and contour preview, the algorithm is very successful in properly classifying species contained in the training library.

The application of digital image processing techniques to the problem of automatic leaf classification began two decades ago and it has been since proceeding in earnest. The technology found some of its earliest applications in industrial agriculture, where the desire was to separate crop species from weed species, allowing for decreased use of pesticides. The problem is complicated in this application by complex backgrounds that make image segmentation difficult, but simplified by foreknowledge of one or two desirable crops amongst unwanted weed species. Color and texture information is often sufficient to make this distinction, as opposed to more general applications, where numerous shape features must be acquired. More recently, several groups have approached the problem of automatic leaf classification. Though the groups often use similar digital morphological features, e.g. rectangularity, sphericity, eccentricity, etc., there is great variation in how these measures are combined and used in classification. *Wu et al*, for instance, used a two-tiered system, eliminating grossly different samples on the basis of eccentricity alone before making a finer judgment based on the combination of the cancroids-contour distance (CCD) curve, ACH, and eccentricity. *Du et al.* used roughly a dozen morphological features and moments and defined a classification method called the move median centers (MMC) hyper sphere classifier, achieving a correct classification rate of over 90%. *Wu et al.*, on the other hand, achieved similar results employing a probabilistic neural network. Another approach looks at a polygonal approximation of the leaf's shape.

Strategies of using near infrared spectroscopy (NIR) to classify plant leaves are as follows: firstly, same number but different types of plant leaves were measured in the laboratory by using the Nexus-870 Fourier transform infrared spectroscopy, then wavelet analysis and Blind Sources Separation (BSS) were used to process the sample data, finally, BP neural network algorithm was applied to classify different kinds of plant leaves. Figure shows the steps taken for the classification of plant leaves. The first step of the experiment is to collect Camphor and Maple leaves, 75 pieces for each. The total pieces are divided into two sets: training set and test set. As is shown in Table, Camphor leaf is marked as Leaf A, Maple leaf is marked as Leaf B, and each 50 pieces of both kinds are treated as training samples, and the rest 25 pieces of each are test samples.

2.ORACLE DATABASE

An Oracle **database** is a collection of data treated as a unit. The purpose of adatabase is to store and retrieve related information. A database server is the key tosolving the problems of information management. In general, a server reliablymanages a large amount of data in a multiuser environment so that many users canconcurrently access the same data. All this is accomplished while delivering highperformance. A database server also prevents unauthorized access and providesufficient solutions for failure recovery.The database has logical structuresand physical structures. Because the physical and logical structures are separate, the physical storage of data can be managed without affecting the access to logical storage structures.

3.PRESENT SYSTEM

Leaf shape description is the key problem in leaf identification. Up to now, many shape features have been extracted to describe the leaf shape. But there is no proper application to classify the leaf after capturing its image and distinguishing its attributes yet.

4. PROPOSED SYSTEM

In the proposed system, the application facilitates user to provide the image of the leaf as the input. The system applies algorithm to derive vital parameters related to the properties of the leaf. It then compares these parameters with the ones stored against a leaf entry in the database.On successful match of the parameters, the application displays information related to that particular leaf to the user for his review.

Likewise, the system also facilitates the user to report false reports generated by the application so as to sharpen its results in future

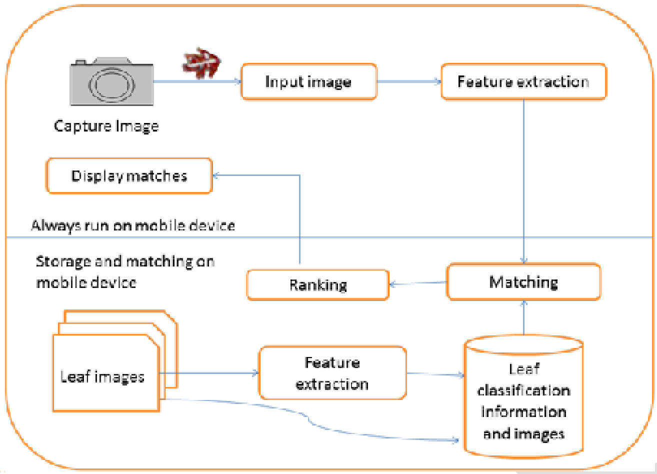


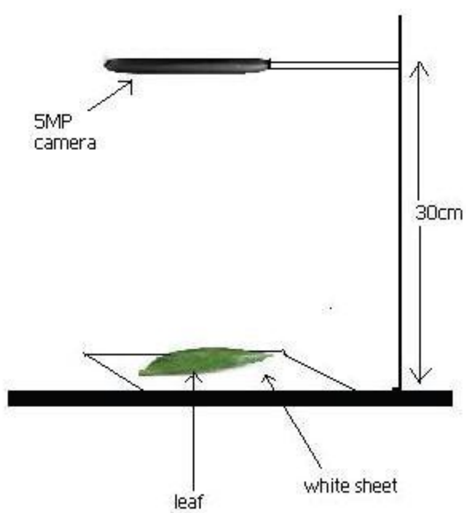
Figure 4.1: System Architecture

4.1 Training

In this module, necessary input is fed to the system in the form of images of the leaf. The system applies necessary steps to extract values for vital parameters from the image. This image along with these parameters with their values and other essential information is stored in the database. These functions are performed by the admin from his login screen

4.2 Image Acquisition

This module is a part of the app installed on the android mobile. This module is initiated whenever the user intends to discover details about any leaf. Using this module, the user can submit the image of the leaf for its identification by the system. This module captures the image and sends it to the central server for processing.



4.3 Preprocessing

Using this module, the image captured from the user's mobile is subjected to necessary preprocessing. In this method, the image is converted into a standard binary format from other format like color or grayscale. This preprocessed binary image is then subjected to identification wherein the vital parameters of the leaf are extracted for its comparison.

4.4 Identification

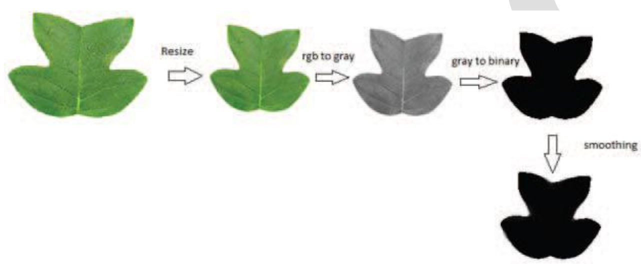
In this process, the vital parameters of the leaf are extracted for its comparison with the ones stored in the database. The algorithm is applied on the preprocessed image for its comparison. The image which has maximum of its characteristics matched with the ones stored in the database is displayed to the user for viewing further details about the leaf

4.5 Improvement by feedback

The system facilitates user with an option to improve the results generated by it and reporting whether they were as per the expectation and reality. Based on the feedback submitted by the user, the system trains and improvises itself further.

4.6 Share

The system also gives the user an option to share the results with other members around him



Algorithm Steps:

1. **Image Acquisition**
   a. Image capture from phone
2. **Pre-processing**
   a. Color-grayscale image to binary image
3. **Morphological Feature Extraction**
   a. Centroid-contour Distance Curve
   b. Aspect Ratio (AR)
   c. Rectangularity (R)
   d. Convex Area Ratio (CAR)
   e. Convex Perimeter Ratio (CPR)
   f. Sphericity (S)
   g. Circularity (C)
   h. Eccentricity (E)
   i. Form Factor (FF)
   j. Regional Moments of Inertia (RMI)

The application of digital image processing techniques to the problem of automatic leaf classification began two decades ago and it has been since proceeding in earnest. In 1989, Petry and Kuhbauch were the first to extract digital morphological features for use with identification models. The technology found some of its earliest applications in industrial agriculture, where the desire was to separate crop species from weed species, allowing for decreased use of pesticides. The problem is complicated in this application by complex backgrounds that make image segmentation difficult, but simplified by foreknowledge of one or two desirable crops amongst unwanted weed species. Color and texture information is often sufficient to make this distinction, as opposed to more general applications, where numerous shape features must be acquired. More recently, several groups have approached the problem of automatic leaf classification. Though the groups often use similar digital morphological features, e.g. rectangularity, sphericity, eccentricity, etc., there is great variation in how these measures are combined and used in classification. *Wu et al*, for instance, used a two-tiered system, eliminating grossly different samples on the basis of eccentricity alone before making a finer judgment based on the combination of the cancroids-contour distance (CCD) curve, ACH, and eccentricity. *Du et al.* used roughly a dozen morphological features and moments and defined a classification method called the move median centers (MMC) hyper sphere classifier, achieving a correct classification rate of over 90%. *Wu et al.*, on the other hand, achieved similar results employing a probabilistic neural network. Another approach looks at a polygonal approximation of the leaf's shape.

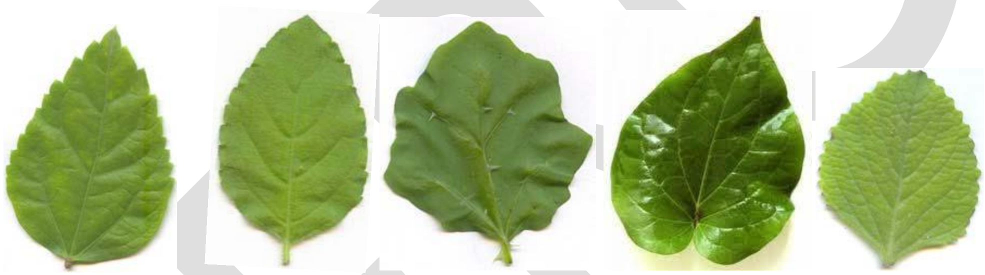
Strategies of using near infrared spectroscopy (NIR) to classify plant leaves are as follows: firstly, same number but different types of plant leaves were measured in the laboratory by using the Nexus-870 Fourier transform infrared spectroscopy, then wavelet analysis and Blind Sources Separation (BSS) were used to process the sample data, finally, BP neural network algorithm was applied to classify different kinds of plant leaves. Figure shows the steps taken for the classification of plant leaves. The first step of the experiment is to collect Camphor and Maple leaves, 75 pieces for each. The total pieces are divided into two sets: training set and test set. As is shown in Table, Camphor leaf is marked as Leaf A, Maple leaf is marked as Leaf B, and each 50 pieces of both kinds are treated as training samples, and the rest 25 pieces of each are test samples.

Many papers have been presented in International Journals and conferences; researchers have worked on hierarchical, neural networks and machine learning methods. The work on classification of leaves started in the early 20th century. The earlier dataset was the laboriously collected set of leaves, gathered by those who were working on it. Chrysanthemum, for instance, was the dataset created and used in involved user interaction to differentiate the 12 varieties, of the same kind that included self-intersections. The Flavia dataset was originally created by Wu et al. It is a collection of 32 different species, a total of 1907 leaves. He extracted 12 leaf features and featured them into 5 principle variables. This was the input vector to the Probabilistic Neural Network (PNN) that was trained by 1800 leaves. It earned an accuracy of more than 90%. This flavia dataset, now available as the standard dataset is accessible to all the researchers to work on. Initially the works were carried out using neural networks.

Tzionas et al., in implemented an artificial vision system that extracted specific geometrical and morphological features. Using a novel feature selection approach, a subset of significant image features was identified. A feed forward neural network was employed to perform the main classification task and that was invariant to size and orientation. It could successfully operate even with the deformed leaves. It achieved a considerable high classification ratio of 99%. Further, Lin and Peng in attempted to realize a computer automatic classification for 30 broad leaved plants in a more convenient, rapid and efficient manner using PNN achieving 98.3% of accuracy. Kadir et al., in also proposed a method for leaf classification which incorporates shape and vein, color and texture features and used PNN as a classifier. The experimental result gives an accuracy of 93.75% when tested on Flavia dataset. Later with the advancement of Machine Learning, researches were conducted comparing its techniques with the neural network approaches. Anami et al., used Support Vector Machine based on color and texture feature and neural network classifier to identify and classify images of medicinal plants such as herbs, shrubs and trees. The edge and color descriptors have low dimension and are effective, simple and rotation -invariant. Singh et al., in carried out the experiment on 50 leaf samples of 32 different classes of varying shapes; using three different techniques; Support Vector Machine (SVM), Fourier Moments and PNN based on the leaf shape and achieved an accuracy of 96%, 62% and 91% respectively. ArunPriya et al., in compared SVM and k-NN classifier on flavia and a real dataset consisting of 15 tree classes and claimed SVM to be a better classifier achieving more than 94% of accuracy for both the dataset. Kumar et al., conducted a survey on different classification techniques; k-Nearest Neighbor classifier, Neural Network, Genetic Algorithm, SVM and Principal Component Analysis and listed their advantages and disadvantages. The drawbacks of SVM are that it is a binary classifier, training is slow, and it is difficult to understand structure of algorithm. It also has limitation with speed and size, both in training and testing. Owing to the drawbacks of SVM, Valliammal and Geethalakshmi in worked on automatic recognition system taking a total of 500 plant and flower images for identifying and recognizing them to their respected categories using Preferential Image Segmentation which is invariant to translation, rotation and scale transformations. They further extended their work towards leaf classification and recognition using Hybrid image segmentation that combines threshold and H maxima transformations.

This method extracted more accurate values of the leaf and involved minimum computational time in comparison. A recent paper by Kaur and Kaur in used neural network based LM algorithm to train the classifier and carried out the experiment on 12 kinds of leaves which yielded an accuracy of 97.9 %. The leaf classification is of great importance in the field of medicine. S.E Kumar in conducted an experimental analysis with a few medicinal plant species such as Hibiscus, Betel, Ocimum, Leucas, Vinca etc. and proved that the method of identification based on leaf features such as area, color histogram and edge histogram is an efficient attempt.

The leaves of five different plants namely Indian borage (Karpoora**V**alli), Hibiscus rosa-sinensis (**H**ibiscus),Ocimumtenuiflorum (**T**ulasi), Solanumtrilobatum (**D**huthuvalai) and Piper betel (**B**etel) were collected, since they are used in most of the traditional herbal cough syrups in India. The images of these plant leaves are shown in the Figure 1. and Figure 2. as fresh and dried ones. A large number of samples were collected and based on their physical well being, few samples were taken for analysis.



CONCLUSION

Our undertaken project is thus going to be useful for farmers ,horticulturists and to trekkers by providing an useful plant leaf identification system and will eventually identify their diseases.

ACKNOWLEDGEMENTS

We would like to sincerely thank Prof.S.P.Godse, our guide from Sinhgad Academy of Engineeringfor his support and encouragement.

REFERENCES:

[1] Z. Wang, Z. Chi, and D. Feng, “Shape based leaf image retrieval, ” *IEE Proc. Vis. Image Signal Process., 150 (1),* pp. 34-43, 2003.

[2] J.X. Du, D.S. Huang, X.F. Wang, and X. Gu, “Computer-aided plant species identification (CAPSI) based on leaf shape matching technique, ” *Trans. Inst. Measure. Control, 28 (3),* pp. 275-284 , 2006.

[3] F. Mokhtarian and S. Abbasi, "Matching shapes with selfintersection: application to leaf classification, " *IEEE Trans. Image Process., 13 (5)*, pp. 653-661 , 2004.
[4] C.L. Lee and S.Y. Chen, "Classification for leaf images, " *Proc. 16th IPPR Conf. Comput. Vision Graphics Image , Process, pp. 355–362, 2003.*
[5] J.X. Du, X.F. Wang and G.J. Zhang, "Leaf shape based plant species recognition, " *Applied Mathematics and Computation, 185, pp. 883–893, 2007.*
[6] X.F. Wang, D.S. Huang, J.X. Du, H. Xu and L. Heutte, "Classification of plant leaf images with complicated background, " *Applied Mathematics and Computation, 205, pp. 916–926, 2008.*
[7] S. Abbasi, F. Mokhtarian and J. Kittler. "Reliable classification of chrysanthemum leaves through curvature scale space, " *ICSSRC97*, pp. 284–295, 1997.
[8] F. Mokhtarian, S. Abbasi and J. Kittler, "Efficient and robust retrieval by shape content through curvature scale space, ".*Proceedings of the International Workshop Image DataBases and MultiMedia Search, pp. 35–42, 1996.*
[9] F. Mokhtarian and A.K. Mackworth, "A theory of multiscale curvature based shape representation for planar curves, " *IEEE Trans. Pattern Anal. Machine Intell., 14, pp. 789-805, 1992*