ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Identification of Multiple Leaves and Retrieval of its Medicinal Value

Nisha Jenifer Roche[1], Dr.Jharna Majumdar[2], Shilpa Ankalaki[3], Santhosh Kumar K L[4]

Department of CSE (PG) , Nitte Meenakshi Institute of Technology, Bangalore ,India[1]

Professor& Head, Department of CSE (PG) , Nitte Meenakshi Institute of Technology, Bangalore, India[2]

Asst. Prof. Department of CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India[3]

Asst. Prof. Department of CSE (PG), Nitte Meenakshi Institute of Technology, Bangalore, India[4]

ABSTRACT: Earth has a large part of flora diversity. However globalization has adversely affected our environment and much of the plants are on the verge of extinction. The chemical compounds present in plants works in a similar manner to that of man-made drugs. As this field is growing, it becomes utmost important to keep the database of plants updated so that identification of plants can be done smoothly. This paper proposes an efficient method to detect multiple leaves called the Component Labeling Method. Shape features which are invariant to translation, rotation and scaling are considered. In addition to the existing basic shape features new features which are also invariant have been considered. The features in the feature database are grouped by using an efficient clustering method called CURE and finally the leaves are identified. Accuracy of the leaves identified using basic features and the new features are also shown.

KEYWORDS: Component Labeling Method, translation, rotation and scaling, CURE Clustering Method

I. INTRODUCTION

Plants play a vital role in our environment. They purify the air and also give food and shelter to animal species. Apart from this they are also used for economical purposes[11].

Earth has a large part of flora diversity. Much of these plants have medicinal value. These plants have been identified and used throughout history. The chemical compounds in the plants work similarly to the chemical compounds present in drugs created by man[11]. Hence herbal medicine does not differ much from conventional drugs. The herbal medicines are very safe as they do not have any side effects. They are also available at a very low cost. Hence most people prefer herbal medicines. As this field is growing there is a high need to identify plants that can be used as medicine. Day to day we here names of new diseases. Creating a new drug is definitely time consuming. Hence if an alternative to these drugs can be found out which doesn't have any side effect is definitely beneficial. With the urbanization a lot of plants are being destroyed and hence the plant species are on the verge of extinction. Thus building a leaf database for quick retrieval of information has become important.

A plant can be identified by its leaves, flowers and fruits. But the leaves play a prominent role in identification because they are almost available throughout the year. A leaf can be identified by its shape features[2]. Different leaves have different shape features. Using these features plant leaf can be identified. Digital image processing is a technique which uses high level algorithms to perform image processing. Thus, the image of the leaf can be processed by image processing techniques and the leaf can be identified.

In recent years numbers of pattern recognition techniques and computer methodologies have been applied for automatically categorizing plant species [9].Several methods have been proposed for identifying single leaf [8-11]. The proposed method can identify both single and multiple leaves when given as a query using shape features. Color and Texture features of the leaves are not considered because as the plant ages its color and texture might change. Thus using these features would not be realistic [11].

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

II. METHODOLOGY

In this paper a methodology has been proposed which will learn the shape features of the leaf and stores it in a database and finally identifies it when given as a query. The proposed methodology consists of 2 main phases

1) Learning Phase
2) Identification Phase

Figure 1 shows the overall system diagram of the proposed method.

Learning Phase: This is the phase where the proposed system learns about its database. In this phase the unknown leaves are collected and stored in the database. Features which are invariant to translation, rotation and scaling are extracted and are stored in the database. These features are then clustered such that features that are most similar are grouped together. Learning phase involves the following steps "data collection, pre-processing, feature extraction, feature normalization and Clustering Algorithms".

Identification Phase: This is the phase where the proposed system identifies the query leaves by comparing their features with the features in the database and retrieves those leaves whose features match exactly or the nearest matched leaves are retrieved.

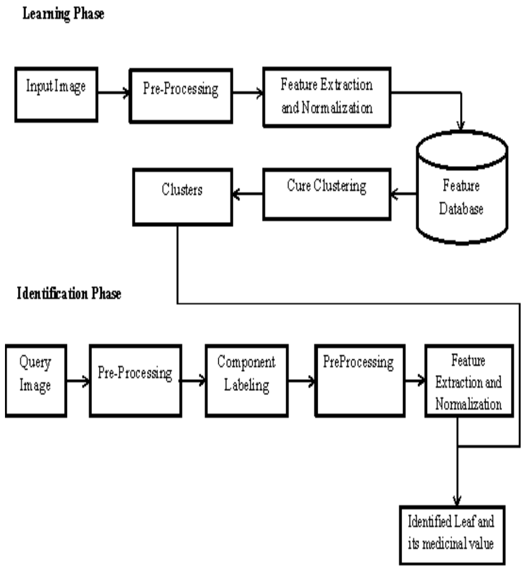


Fig 1: Overall System Diagram

A. *Learning Phase*

The Proposed System for Identification of multiple leaves consists of a camera fitted with a clamp on a table top and is connected to the computer using a suitable interface.A light source is also used to properly illuminate the area.

1)Data Collection for Multiple Leaves Identification and Pre-Processing

The healthy leaves having medicinal value were captured using any photographic equipment. Next the invariant shape features were calculated. All the shape features are explained in the next section

2) Feature Selection

a) List of Morphological Features of the Leaf

- **Filled Leaf Area [2]**: This feature gives the amount of space occupied by the leaf.
- **External Leaf Perimeter [2]**: This feature is obtained by counting only the pixels present at the boundary of the leaf.
- **Leaf Fiber Length [10]:** Is the length of the middle line of the curved feature. Leaf Fiber Length = (Filled Leaf Area/2) - 2(Filled Leaf Area/ External Leaf Perimeter)
- **Leaf Curl [10]:** This feature is a ratio of Leaf maximum Length to Leaf Fiber Length.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

- **Half Width Ratio [8]:** The ratio of width at half of the maximum width to maximum length of the leaf.
- **Leaf Compactness [2]:** The ratio of sum of all pixels in the leaf to the square of sum of boundary pixels.
- **Form Factor of the Leaf [2]:** Is a measure of leaf area and leaf perimeter.
- **Half Leaf Ratio of a Leaf [2]:** This feature is obtained by first finding the smallest box that can enclose the shape. This box is then divided into half. Then the upper leaf area is calculated. Similarly the lower leaf area is also calculated. The ratio of this upper to lower area gives the half leaf ratio of the leaf. Fig.2(c) shows the leaf aspect ratio.
- **Leaf Aspect Ratio [2]:** The ratio of maximum length (MaxL) to maximum width (MaxW). Fig.2(a) shows the leaf aspect ratio
- **Rectangularity of the Leaf [2]:** Ratio of filled area of the leaf to the filled area of the smallest box that can enclose a shape.
- **Leaf Eccentricity [2]:** The ratio of the two principal axes.

$$E = \frac{\text{Minor Principal Axis}}{\text{Major Principal Axis}}$$

- **LeafCircularity[2]:** Is a feature which helps in deciding how much a shape is similar to a circle.

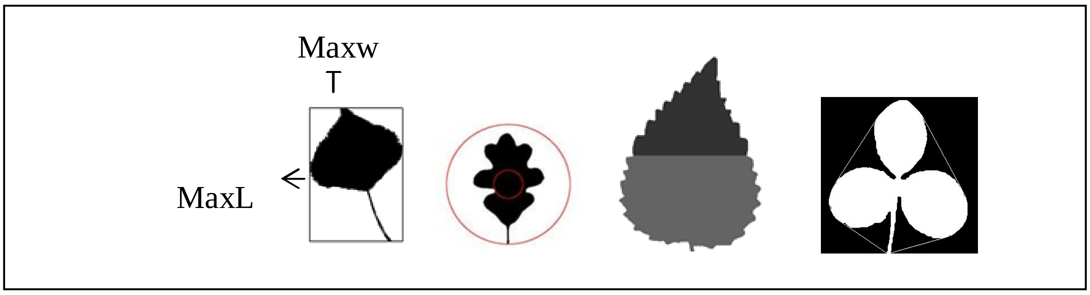
$$C = \frac{4 \pi A}{P^2}$$

- **LeafSphericity[2]:** It is the ratio of smallest circle that can be inscribed in the shape to the largest circle that can be inscribed in the shape. Fig.2(b) shows the leaf sphericity.
- **Perimeter Ratio[8]:** The ratio of sum of boundary pixels to the sum of maximum length and maximum width
- **Elongation [2]:** This feature is given by the square root of minor to major principal axis.

b) Leaf Convex Hull [2]

"A Convex Hull of a leaf is the smallest polygon that contains all the points of the object .i.e. if any two points of the object are tried to join then they must lie within the polygon". Fig. 2(d) shows an example of Convex Hull.

- **Waviness Factor [8]:** "The ratio of perimeter of the convex to the perimeter of the object".
- **Solidity[13]:** (Area of the object) / Area of the convex hull



c) Moment Invariant Features

The Invariant Moments are widely applied because of their unique property of "translation, rotation and scale invariance" [3].In our proposed work Hu moments were used. Moment Invariants were first introduced by Hu. He proposed 6 orthogonal invariant moments and one skew invariant moment [3] based on algebraic moments. A two dimensional central moment is given by

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)\,dx\,dy \qquad p,q = 0,1,2,\cdots$$

The 7 invariant moments introduced by Hu are as follows

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

$$\phi_1 = \eta_{20} + \eta_{02}$$
$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \mu_{03})^2$$
$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \mu_{03})^2$$
$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$
$$\phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$
$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

Using regular shape descriptors may not be able to identify a leaf accurately. Hence using moment invariant features helps in increasing the accuracy of identification.

3)Data Normalization

“Feature Normalization is a process of normalizing the features so that they fall within a particular range”. The raw data has different ranges of values. Normalizing these values are essential because most of the clustering and classification algorithms compute distance between two features. If one of these features has a wide range of values then the distance will be wrongly calculated. In the proposed work normalization is done by finding the minimum (Xmin) and maximum (Xmax) feature in each row and then subtracting it from the current feature Xi where i=1, 2…22

$$D = \frac{(Xi-Xmin)}{(Xmax-Xmin)}$$

4) Cure Clustering [6]

Clustering is an unsupervised machine learning technique. In our proposed work clustering is done so that leaves in a cluster are much similar to each other when compared to leaves in different clusters.

Algorithm-Cure Clustering

Input: Feature Database S, No. of Clusters k
Output: Required number of clusters
Step 1: Read features from the Feature Database S into an array
Step 2: The features are then inserted into the heap in the increasing order of their distances.
Step 3: The topmost element from the heap is extracted since it has the minimum distance. This element gives the features that are to be merged.
Step 4: These features are then merged by taking the average of the two features. This is called as wmean. The new representative point is chosen by using the following formula
Wrep = wrepU{p+ alpha*(wmean - p)} where alpha =0.3
Step 5: Delete the features that were used for merging from the heap.
Step 6: Insert the merged feature Wrep into the heap.
Step 7: The heap is again sorted so that the features are in ascending order and Step 3 is repeated.
These steps are continued till the required number of clusters are obtained.

B. Identification Phase

In this phase, leaves which are to be identified are placed in from of the camera and its image is acquired, it is then pre-processed and then component labeling method is used to detect multiple leaves, features of each leaf from the query image is extracted, they are then normalized and matched with the database based on the distance measure and the leaves are identified. The query input can be a single leaf or multiple leaves.

1) Detecting Multiple Objects -Component Labeling[5]

Once the query image is pre-processed Component Labeling method (Blob Extraction) is used to detect multiple leaves.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Algorithm- Component Labeling
Input: Binary Image
Output: Components

First scan
Step1: Initialize an array Eq_label with labels [0-255] and curr_ label =0
For each foreground pixel encountered
Step 2:check its top and left neighbor
Step 3: if the top and left neighbor is a background pixel then
Step 4:curr_label++;
Step 5: Else if the top or the left neighbor is a background pixel then
Step 6: Assign the label of the pixel which is not background
Step7: Else if label of the top pixel is not equal to the label of the left pixel and both the pixels are not background pixels then
Step 8: Scan the Eq_label array from the beginning and replace the label of the top pixel by the label of the left pixel in the Eq_label array

Second scan
Step9: For each foreground pixel encountered replace it with the label in the Eq_label array.

III. EXPERIMENTAL RESULTS

A set of 28 leaves have been captured. The captured leaves should have white background. These leaves were then converted to a size of 256x256 and were stored in RAW format along with their names and their medicinal values for further processing (Fig3).After that pre-processing steps were carried out in which the image was first converted to grey scale(Fig 4) and then to binarized image(Fig 5).

A. *Sample Database*

Leaf 1:Circinatum Leaf 2:Dandelion Leaf 3: Clover Leaf 4: Dock
Fig 3: Example of Leaves Stored in Database

Fig 4: Greyscale Images of Leaves Fig5: Pre-Processed Images of Leaves

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

B. Resultsof Cure Clustering Algorithm

In the proposed work 28 different leaves are used to train the database .The leaf names and their medicinal values are also stored in the database. Fig. 6 shows the results of CURE Clustering Algorithm for k=5 where k is number of clusters

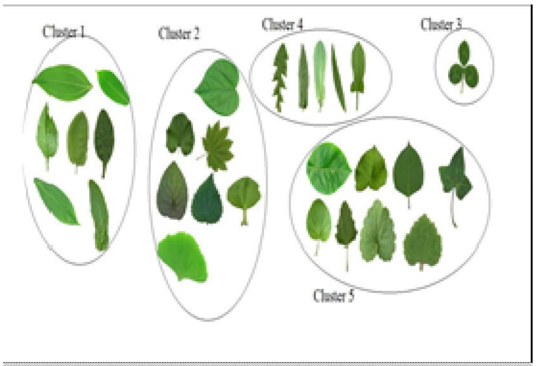


Fig 6.Cure Clustering Results for k=5

C. Results of Component Labeling

The query image consisting of multiple leaves is given as input to the proposed system, After preprocessing Component labeling method is applied to detect number of leaves. Fig 7shows the steps involved to get component labeling output.

Fig 7(a) Input Color Image Fig 7(b) GreyScale Image Fig 7(c) Binary Image

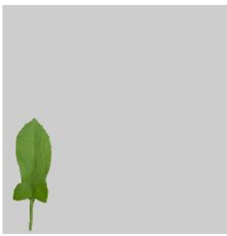


Fig 7(d)

Fig7(e) Fig 7(f) Fig 7(g)

Fig 7(d-g) Results of Component Labeling,The four leaves of one image are detected and put in 4 different buffers.

D.Results of Leaf Identification

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

Figure 8 shows the results of multiple leaf identification. As seen in fig.8 leaves have been identified.Table I shows the medicinal values of the Identified Leaf.

Fig. 8(a)

Fig.8(b) Fig.8(c) Fig. 8(d) Fig
8(e)

Fig.8: (a)Query Image containing 4 Leaves as input image (b)Identified leaf Ivy (c) Identified Leaf Oilseed Rape (d) Identified Leaf Sorrel Sheep (e) Identified Leaf HorseChestnut

TABLE I. MEDICINAL VALUE OF THE IDENTFIED LEAF

Leaf Name	Medicinal Value
Ivy	Treatment of Respiratory Problems
Oilseed Rape	Treatment of Chronic Cough
Sorrel Sheep	Treatment of Cancer
Horsechestnut	Treatment of Arthritis

E. Comparison of accuracy of leaf identification

Adding relevant features definitely increases the accuracy of leaf identification. In our proposed work leaf identification was done when both single and multiple leaves where given as query. The initially identification was done using15 features. Later 6 more features were added which are Leaf Fiber Length, Leaf Curl, Half Width Ratio, Waviness Factor, Perimeter Ratio, Elongation which increased the accuracy of identification. Table II and Table IIIshows the accuracy of identification for 15 and 22 features for both single and multiple leaves

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

TABLE II.Analysis using 15 Features

Total leaves in the database	Type of Input	Correctly Identified Instances	Wrongly Identified Instance	Accuracy
28	Single Leaf	28	0	100%
28	Multiple Leaves	17	11	60.71%

TABLE III Analysis using 22 Features

Total Leaves in the database	Type of Input	Correctly Identified Instances	Wrongly Identified Instance	Accuracy
28	Single Leaf	28	0	100%
28	Multiple Leaves	17	11	60.71%

IV. Conclusion and Future Work

ThePlants play a vital role in our lives. However globalization has adversely affected our environment. As a result of this flora diversity is at the verge of extinction. Hence there is a need to build a database that can be used to identify and protect the plants that can be used for some purpose.

The objective of this project was to identify multiple leaves and to retrieve their medicinal value. In order to detect multiple leaves a method called as Component Labeling also known as Blob Extraction was used. The database contained a set of 28 leaves along with its features and medicinal value. In addition to the invariant features mentioned in [1] six new features such as Waviness Factor, Leaf Curl, Leaf Fiber Length, Perimeter Ratio of Diameter, Branch Length Ratio and Perimeter Ratio were implemented. All the new features implemented were found to be scale invariant.

In the present work CURE Clustering Algorithm [6]was used to cluster the leaves based on their features. In future the same work can be done using other Data Mining Clustering Algorithms like BIRCH, OPTICS, Chameleon and a comparison among these clustering methods can be made. The proposed method to detect multiple leaves or objects is useful when the leaves or objects are disjoint. In the future this work can be extended to detect connected objects.

REFERENCES

1. https://en.wikipedia.org/wiki/Otsu's_method.
2. Mingqiang Yang, KidiyoKpalma, Joseph Ronsin. “ A Survey of Shape Feature Extraction Techniques.” Peng-YengYin. Pattern Recognition, IN-TECH, pp.43-90, 2008.
3. Zhihu Huang, JinsongLeng “Analysis of Hu's Moment Invariants on Image Scaling and Rotation” ,Edith Cowan University,2010,IEEE
4. A.Kannan, Dr.V.Mohan,Dr.N.Anbazhagan, “Image Clustering and Retrieval using Image Mining Techniques” IEEE International Conference on Computational Intelligence and Computing Research,2010
5. Luigi Di Stefano, Andrea Bulgarelli , “A Simple and Efficient Connected Components Labeling Algorithm”, Italy
6. SudiptoGuha, Rajeev Rastogi, Kyuseok Shim, “ CURE : An Efficient Clustering Algorithm for Large Databases”
7. Aastha Joshi ,RajneetKaur, “A Review: Comparative Study of Various Clustering Techniques in Data Mining” ,International Journal of Advanced Research in Computer Science and Software Engineering,March 2013
8. Hazra, K.Deb, S. Kundu , P.Hazra , “Shape Oriented Feature Selection for Tomato Plant Identification” International Journal of Computer Applications Technology and Research Volume 2– Issue 4, 449 - 454, 2013, ISSN: 2319–8656

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 7, July 2015

9. Dr.AnjaliS.Bhalchandra, AjinkyaP.Nilawar, "Image Retrieval Using Shape Measures and BTC " International Journal of Advanced Research in Computer Science and Electronics Engineering Volume 1, Issue 3 , May 2012
10. "Phytolith Systematics" By Society for Archaeological Sciences
11. Pallavi P1, V.S Veena Devi "Leaf Recognition Based on Feature Extraction and Zernike Moments" International Journal of Innovative Research in Computer and Communication Engineering, Vol.2, Special Issue 2, May 2014
12. P. S. Hiremath , JagadeeshPujari ,"Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement"
13. J. Du, X. Wang, and G. Zhang, "Leaf shape based plant species recognition," Applied Mathematics and Computation, vol. 185-2, pp. 883-893, February 2007.
14. David Knight, James Painte, Matthew Potter, "Automatic Plant Leaf Classification for a Mobile Field Guide".

BIOGRAPHY

Ms.Nisha Jenifer Roche received B.E in Information Science and Engineering from P.A College ofEngineering and Management,Mangalore, Karnataka, India under Visvesvaraya Technological University (VTU) in 2013 and is currently pursuing the P.G (M.Tech) degree in Computer Science and Engineering at NitteMeenakshi Institute of Technology,Bangalore,Karnataka, India under VTU. Her current research interests include image processing.

Dr.JharnaMajumdar, presently the Dean of R&D and the HOD of CSE (PG). She served DRDO from 1990 to 2007 and retired as Scientist G and Head of Aerial Image Exploitation Division, Aeronautical Development Establishment (DRDO), Bangalore, India.Dr.Majumdar received B.Tech (Hons.) in Electronics and Electrical Engineering and Post Graduate in Computer Technology from Indian Institute of Technology Kharagpur in 1969 and 1970 respectively. She received her PhD (Electrical Engineering) in 1980

Ms.ShilpaAnkalaki, presently working as Asst. Professor in NitteMeenakshiInstitute of Technology, Bangalore, Karnataka, India. She received Masters of Technology from NitteMeenakshi Institute of Technology under VTU, Bangalore, Karnataka, India. Her area of interests are Digital Image Processing, Big Data, Pattern Recognition.

Mr.Santhosh Kumar K L,presently working as Asst. Professor in NitteMeenakshi Institute of Technology, Bangalore,Karnataka, India. He received Masters of Technology from NitteMeenakshi Institute of Technology underVTU,Bangalore, Karnataka , India. Hisareas of interests are Digital Image Processing, Video Shot Detection and Summarization, Facial Expression Recognition, Activity Tracking.