RECOGNITION OF PLANTS BY LEAF IMAGE USING NEAREST NEIGHBORHOOD CLASSIFICATION

Ankur Gupta¹ (M. Tech. Scholar), Dr. B. S. Rai² (Professor)
Department of Electronics and Communication Engineering
Madan Mohan Malaviya University of Technology
Gorakhpur, India.

Abstract: This paper presents a simple and computationally good method for plant species recognition using leaf images. Recognition of plant images is one of the research topics of computer vision. The use of shape for recognizing objects has been actively studied since the beginning of object recognition in 1950s. Several authors suggest that object shape is more informative than its appearance properties such as texture and color vary between object instances more than the shape. Initially we have scanned leaf images which are two dimensional in nature and segmented the images by mathematical morphological segmentation and then extracted the high frequency feature of image. For removing the noise, the image has been converted into binary, than complemented and multiplied by filtered image. We quantitatively establish the use of texture for detection various leaf images of same tree that are difficult by other classical methods of image processing. Further we use Nearest Neighborhood classification method to classify plant leaf. In this paper we focuses mainly on image enhancement, image segmentation, high frequency feature extraction, noise remove from background, volume fraction, inverse difference moment, moment invariant and morphological feature such as area convexity.

Keywords: Plant leaf classification, Moment Invariants, Image Processing, PNN, PCA, Texture analysis, neural networks.

I. INTRODUCTION

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 also very close. In addition, plants are important means of livelihood and production of human beings. Ayurveda [17] is considered a form of alternative to allopathic medicine in the world. This system of medicine has a rich history with a number of ayurvedic leaves which can’t be recognized by a human being. Plants can be classified according to the shapes, colours, textures and structures of their leaf, bark, flower, seedling and morph. Nevertheless, if the plant classification is based on only two dimensional images, it is very difficult to study the shapes of flowers, seedling and morph of plants because of their complex three dimensional structures. Plant leaves are two dimensional in nature and hold important features that can be useful for classification of various plant species. Therefore, in this research, the identification of different plants species is based on leaf features. Research on the utilization of moments for object characterization in both in- variant and non-invariant tasks has received considerable attention in recent years [8, 9]. A substantial amount of work has been done on leaf shape based plant classification and recognition. Wu et al. [1] extracted 12 commonly used digital morphological features which were orthogonalized into 5 principal variables using PCA [24]. They used 1800 leaves to classify 32 kinds of plants using probabilistic neural network system [23]. Wang et al. [2] employed centroid contour distance (CCD) curve, eccentricity and angle code histogram (ACH). Fu et al. [3] also used centroid-contour distance curve to represent leaf shapes in which an integrated approach for an ontology-based leaf classification system is proposed. For the leaf contour classification, a scaled CCD code system is proposed to categorize the basic shape and margin type of a leaf by using the similar taxonomy principle adopted by the botanists. Then a trained neural network is employed to recognize the detailed tooth patterns. Du et al. [4] an efficient computer-aided plant species identification (CAPSI) approach is proposed, which is based on plant leaf images using a shape matching technique. Firstly, a Douglas-Peucker approximation algorithm is adapted to the original leaf shapes and a new shape representation is used to form the sequence of invariant attributes. Then a modified dynamic programming (MDP) algorithm for shape matching is proposed for the plant leaf recognition. Finally, the superiority of our proposed method over traditional approaches to plant species identification is demonstrated by experiment. Gu et al. [5] used the result of segmentation of leaf’s skeleton based on the combination of wavelet transform (WT) and Gaussian interpolation. It is a a new approach for leaf recognition also using the classifiers, a nearest neighbour classifier (1-NN), a k -nearest neighbor classifier (k-NN) and a radial basis probabilistic neural network (RBPNN) are used, based on run-length features (RLF) extracted from the skeleton to recognize the leaves. Finally, the effectiveness and efficiency of the proposed method is demonstrated by several experiments. Wang et al. [6] extracted several geometric features like rectangularity, circularity, eccentricity and seven moment invariants for classification. He introduces a method of recognizing leaf images based on shape features using a hyper sphere classifier. Some [1],[3],[7] approaches employed artificial neural network for its fast performance. Others [5],[6] employed k-nearest neighbor (k-NN) classifier to classify plants. Du et al. [7] introduced shape recognition based on
radial basis probabilistic neural network which is trained by orthogonal least square algorithm (OLSA) and optimized by recursive OLSA. Historically, Hu [8] published the first significant paper on the utilization of moment invariants for image analysis and object representation in 1961. Hu’s approach was based on the work of the nineteenth century mathematicians Boole, Cayley and Sylvester on the theory of algebraic forms. Hu’s Uniqueness Theorem states that if \( f(x, y) \) is piecewise continuous and has nonzero values only in the finite part of the \( f(x, y) \) plane, then geometric moments of all orders exist. It can then be shown that the moment set \( \{m_{pq}\} \) is uniquely determined by \( f(x, y) \) and conversely \( f(x,y) \) is uniquely determined by \( \{m_{pq}\} \).

T.H.Resis [9] stated that moment invariant for pattern recognition presented by hu is incorrect. The four moment absolute invariant under general linear transformation is in error. So he presented the revised fundamental theorem and gave the corresponding absolute moment invariant under general linear transformation. Sidhartha Maître [10] consider the change of effect of contrast in an image and modified the seven moment invariant given by M.K.HU which are independent of change of transformation, scale, rotation and contrast also. Dr Dinesh P Mital[11] proposed an unsupervised texture segmentation technique using multi-channel filtering. This simplicity is due to direct result of decomposition of the original image into several filtered images with limited spectral information. Somkait Udomhansakul and Pichet Wongsita [12] proposed a feature extraction approach in medical magnetic resonance imaging (MRI). Jan Fusser [19] proposed a new set of moment invariants with respect to rotation, translation, and scaling suitable for recognition of objects having -fold rotation symmetry. Du and Zhang [13] approach to a new classification method, named as move median centers (MMC) hyper sphere classifier, for the leaf database based on digital morphological feature is proposed. In particular, by comparing with the nearest neighbor (1-NN) and k-NN classifiers, it can be found that the MMC classifier can not only save the storage space but also reduce the classification time. The proposed method is more robust than the one based on contour features since those significant curvature points are hard to find. An uncomplicated and computationally effective technique for plant species recognition by means of leaf image is recommended by Hossain and Amin [20]. A new technique for feature extraction from natural image like plant leaf is developed by Prasad et al. [21] for automated living plant species identification which would be helpful for botanical students to carry out their research for plant species identification. A novel multi-resolution and multidirectional Curvelet transform is executed on sub segmented leaf images to obtain leaf information, precisely in order that the orientation of the object in the image does not taken into account and which also enhance the accuracy rate. Abdul kadir [22] build a foliage plant identification system for 60 kinds of leaves. It was dedicated to handle two or more plants that have similar/same shape but the colour patterns on the leaves were different. In this case, Zernike moments were combined with other features: geometric features, colour moments and gray-level co-occurrence matrix (GLCM).

This paper presents an easy leaf recognition algorithm. Section 2 discusses image preprocessing and acquisition which includes the image enhancement and segmentation and high frequency feature extraction of a leaf images. Section 3 introduces the texture analysis and extraction of feature descriptors to classify leaf images i.e. parametric calculations as shown in Fig. 1 The plant classifier is presented in section 4 and section 5 discusses the result. Section 6 concludes this paper.

II. IMAGE ACQUISITION AND PRE-PROCESSING

A. Image Enhancement and Segmentation

The first thing which comes in our mind is that the margin of leaf which is essential for our preprocessing algorithm. We enhance our images by first converting into gray scale and then make 3*3 windows and move this window from left to right and top to bottom calculates average of window in every move from left to right and top to bottom. If average of window pixel is greater than gray image pixel value \( d(x,y) \) then replace pixel value with average value otherwise value obtained by Gaussian function which is obtained by:

\[
y = y + (y*c) / (1+ (1/2.71828^y)) \tag{1}
\]

The purpose of segmentation is separation of leaf objects from background so that we can properly use the image features. The output of image segmentation is a binary image in which the leaf objects are numerically displayed with 1 and the background with 0. For segmentation we applied binary mathematical morphological algorithm [16],[18] based on shape of image. In morphological operation, opening in which erosion followed by dilation using 8 connected neighborhood has been performed. After applying above operation we get leaves of image shown in Fig. 2 of

![Flowchart of our proposed Algorithm](image-url)
the complete image. Again to get branch we have performed
dilation followed by erosion to segment image. The result is
shown in Fig. 3.

Fig. 2: Original Image (left) Enhanced Image (right)

Fig. 3: Image Segmentation after Opening and Closing

B. High Frequency Feature Extraction
In a leaf image, edges and sharpness of image contribute
significantly to high-frequency content which contains an
internal structure or texture of leaf image which is same for
that plant leaf. A high pass filter yields edge detection in the
spatial domain, because edges contain many high frequency
components. Areas of rather constant gray level consist of
mainly low frequencies and are therefore suppressed. For
extraction of high frequency components of leaf images we
take Fourier transform [14] and shift all low frequency
components to center of the image. Further it is passed
through Butterworth high pass filter and re-transform into the
spatial domain. Then we attenuate low frequency
components in leaf image as shown in the Fig. 4 and Fig. 5
and the pictorial representation of image preprocessing is
given in Fig. 6.

Fig. 4: Original image (left) Butterworth filter (middle)
Filtered image (right)

Fig. 5: Original image (left) filter image with noise (middle)
Filtered image without noise (right)

Fig. 6: A pre-processing example

III. DIFFERENT MORPHOLOGICAL FEATURE
EXTRACTION AND PARAMETRIC CALCULATIONS

A. Smooth Factor
The effect of noises to image area is used to illustrate the
smoothness of leaf image. Smooth factor is the ratio between
area of leaf image smoothed by $5 \times 5$ rectangular averaging
filter and the one smoothed by $2 \times 2$ rectangular averaging
filter.

B. Aspect Ratio
It is the ratio of length L to width W i.e. L/W.

C. Leaf Area
Area is the actual number of pixels in the region. The area of
leaf in a preprocessed image is the number of white or '1'
pixels.

D. Rectangularity
Rectangularity illustrates the similarity between a leaf and a
rectangle. It is defined as $LW/A$, where L represents length,
W denotes the width and A is the leaf area.

E. Circularity
Circularity is ratio involving area of the leaf A and square of
perimeter P of the leaf. It can be defined as $A/P^2$.

F. Eccentricity
A scalar value which specifies the eccentricity of the ellipse
has the same second moments as the region. The eccentricity
is the ratio of the distance between the foci of the ellipse and
its major axis length. The value ranges between 0 and 1.
G. Solidity

Solidity is defined as ratio between the area of the leaf and the area of its convex hull. It is defined as the S (area of leaf/area of convex).

H. Moment Invariant

Moments and functions of moments have been extensively employed as invariant global features of images in pattern recognition. Image moment is a certain particular weighted average (moment) of the image pixel' intensities, or a function of such moments, usually chosen to have some attractive property or interpretation. The idea of using moments in shape recognition gained prominence when Hu, derived a set of seven invariants using algebraic invariants [8]. Here we calculate all seven moment invariants derived by Hu for different leaves of particular plant. In particular, moment functional have attracted great attention due to their mathematical simplicity and numerous physical interpretations. Let \( \{ \mu_n \} \) be a real sequence of numbers and let us define by (2)

\[
\Delta^n \mu_n = \sum_{i=0}^{\infty} (-1)^i \binom{n}{i} \mu_{n-i} + i
\]

(2)

\[
\phi_1 = \mu_{20} + \mu_{02}
\]

\[
\phi_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}
\]

\[
\phi_3 = (\mu_{30} - 3\mu_{12})^2 + 3(\mu_{21} + \mu_{03})^2
\]

\[
\phi_4 = (\mu_{30} - 3\mu_{12})(\mu_{20} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2
\]

\[
\phi_5 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03})
\]

\[
\phi_6 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2
\]

I. Volume Fraction

Volume fraction can be used as a feature descriptor to identify plant. Let \( w \) be the raster grid of pixels and \( \phi \) be the pixel with value one. Then volume fraction is estimated by using (6)

\[
\rho = \frac{v(\phi)}{v(w)}
\]

(6)

Where \( V \) the number of the pixel having value one and \( v(w) \) is the total number of pixels. This feature can be used for distinguishing between various leaf images from different plants.

J. Inverse Difference Moments

An image texture is a set of metrics computed in image processing intended to enumerate the apparent texture of a leaf image. Leaf Image Texture gives information regarding the spatial arrangement of color or intensities in a leaf image or selected region of a leaf image. The recognition of explicit textures in an image is achieved principally by modeling texture as a two dimensional gray level variation. The resultant two dimensional arrays are called as Gray Level Co-occurrence Matrix (GLCM). Inverse difference moment is one of the feature descriptor of GLCM image to identify texture image. This parameter is also called Uniformity. Inverse Difference Moment is also called the "Homogeneity". Mathematically, it can be written as:

\[
\sum_{i,j} P(i, j) \left( \frac{1}{i-j} \right)
\]

(7)

K. Area Convexity

Convexity is any line drawn through the object (and not tangent to an edge or corner) meets its boundary exactly twice and area convexity is related to the geometry of the shape. Area convexity defines that how many times a single line can cut the closed inner or outer geometry of the shape; whether it is the simple closed region or complex closed region. It is the ratio of the perimeter of the convex hull of the sample (Pconvex Hull) over the actual perimeter of the sample (P).
IV. PLANT CLASSIFIER

In our experiment, the seven features that are potentially insensitive to changes in image size, shape are searched and extracted from the images. We used the nearest neighborhood classifier in our study. We measure seven features descriptor of different plant leaves and stored in dynamic matrix M. Further we generate an xls sheet and write this matrix in xls sheet which is stored in our current program directory. After that enter a test image which is also from one of them plant but not that leaf which we have taken. In our experiment we find the Euclidean distance of test image from all stored image into M and search which image has minimum distance from test image that image is our matched image. We perform our experiment on more than 300 leaves of 14 different plants.

Let M is our matrix where we stored features vectored of different images

\[
M = \begin{bmatrix}
m_{11} & m_{12} & \ldots & \ldots & m_{1j} \\
m_{21} & m_{22} & \ldots & \ldots & m_{2j} \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
m_{m1} & m_{m2} & \ldots & \ldots & m_{mj}
\end{bmatrix}
\]

Where \( m_{ij} \) is \( j^{th} \) feature vector of \( i^{th} \) image

Parameter of test image is measure and is represented by

\[
A = [a_{11} \ a_{12} \ \ldots \ \ldots \ a_{ij}]
\]

Where a\( _{ij} \) is \( j^{th} \) feature vector of test image

For recognition of leaves Euclidean distance are measure from the stored parameter in the xls sheet of program directory with the test image and minimum distance show the possible matching with test image

\[
R(1) = (a_{11} - m_{11})^2 + (a_{12} - m_{12})^2 + (a_{13} - m_{13})^2 + (a_{14} - m_{14})^2 + (a_{15} - m_{15})^2
\]

\[
R(2) = (a_{11} - m_{21})^2 + (a_{12} - m_{22})^2 + (a_{13} - m_{23})^2 + (a_{14} - m_{24})^2 + (a_{15} - m_{25})^2
\]

\[
R(3) = (a_{11} - m_{31})^2 + (a_{12} - m_{32})^2 + (a_{13} - m_{33})^2 + (a_{14} - m_{34})^2 + (a_{15} - m_{35})^2
\]

\[
R(4) = (a_{11} - m_{41})^2 + (a_{12} - m_{42})^2 + (a_{13} - m_{43})^2 + (a_{14} - m_{44})^2 + (a_{15} - m_{45})^2
\]

\[
R(5) = (a_{11} - m_{51})^2 + (a_{12} - m_{52})^2 + (a_{13} - m_{53})^2 + (a_{14} - m_{54})^2 + (a_{15} - m_{55})^2
\]

\[
R(6) = (a_{11} - m_{61})^2 + (a_{12} - m_{62})^2 + (a_{13} - m_{63})^2 + (a_{14} - m_{64})^2 + (a_{15} - m_{65})^2
\]

V. RESULT AND DISCUSSION

We prepare our database for the experimental use. The database contains various leaves with various shapes, colors and size. Experiment was done with these different leaves of different classes and tested in our classifier. First we read different leaf images of plant and finally match the new input image with our previous class or tree. We checked the accuracy that it matches the actual image or not. For all the 320 leaves of different 14 plants taken, which are completely different in their shape, color. The various parameters like volume fraction are found to be from 0.30528 to 0.38267 which is within 10 percent variation as shown in Fig. 7. Like all previous image feature descriptor parameter, inverse difference moment also have same value or variation within 10 percent as shown in Fig 8. Hence, this feature can be distinguishing plant leaf images. We calculate area convexity of more than 20 leaf of one plant shown in Fig. 9 which is different in size and color and value of area convexity of each leaf of same plant is coming out to be same or variation within 5 percent.
All the experiments are programmed by Matlab[15] and run on Intel(R) core i3 with the clock of 2.40 GHz and the RAM of 2GB under windows 7 environment. The database subset of some leaf images are shown in Fig. 10 and the experimental work done can be understand by Fig. 11, Fig. 12 and Fig. 13. The Details about the Leaf Numbers of Different Types of Plants are given in Table 1.
The following result will be displayed on the MATLAB command window, like this:

(0.002)

(minimum distance of all calculated data base image into test)

image) sawq = 0.002 name2 = chestnut_leaf.jpg path2 = G:\study\Thesis\A Leaf Recognition Algo\Leaves_IMAGE\)

Table. 1: Details about the Leaf Numbers of Different Types of Plants

<table>
<thead>
<tr>
<th>Class</th>
<th>Common Name</th>
<th>No. of leaf samples</th>
<th>No. of incorrect recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1</td>
<td>Chestnut leaf</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Class 2</td>
<td>Golden rain tree</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Class 3</td>
<td>Trident maple</td>
<td>23</td>
<td>4</td>
</tr>
<tr>
<td>Class 4</td>
<td>Chinese redbud</td>
<td>25</td>
<td>3</td>
</tr>
<tr>
<td>Class 5</td>
<td>Horse chestnut</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>Class 6</td>
<td>Bamboo</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>Class 7</td>
<td>Rose</td>
<td>20</td>
<td>3</td>
</tr>
<tr>
<td>Class 8</td>
<td>Eenchriinigherfslad</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Class 9</td>
<td>Autumn leaf</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Class 10</td>
<td>Pipe</td>
<td>30</td>
<td>2</td>
</tr>
<tr>
<td>Class 11</td>
<td>Golden Maple Leaf</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Class 12</td>
<td>Japan Arrowwood</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Class 13</td>
<td>Castor aralia</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>Class 14</td>
<td>Canadian poplar</td>
<td>25</td>
<td>3</td>
</tr>
</tbody>
</table>

The experiment is designed to illustrate the performance of two feature extraction methods, Gray Level Co-occurrence Matrix (GLCM) and Moment Invariant. GLCM method in leaf recognition for the degrees 0° and 90° gave the same accuracy and same result. Here the poor result is in the 45° degree. Because any changes in the neighboring distance or the neighboring degree it will change the value of extracted texture feature. The GLCM method is very sensitive for the any changes in the images such rotation, scale and etc. Image processing techniques are used for extracting the morphological parameters that are having some significance and effect on the classification of the leaves. Out of total sample of 320 leaves of 14 kinds of plants 293 were classified and 27 were misclassified, that is, a recognition accuracy of 91.5%.

VI. Conclusion

In this paper we propose an automated system for plant identification using shape features of their leaves. It has been found that four parameters that are area convexity, volume fraction, moment invariant, inverse difference moment, provide better results. We conclude that it is a feasible alternative for classifying structurally complex images. They offer exceptional invariance features and reveal enhanced performance than other moment based solutions. The experimental results explained that the proposed method is effective. However, some other works will be explored to obtain better performance. Our future research works will include how to classify the leaves with deficiencies and combine adaptive neural networks to increase the more correct recognition rate.

VII. ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their comments and suggestions to improve this paper. We also wish to express our sincere thanks to all faculty members of the Electronics & Communication Engineering Department for their Cooperation and valuable suggestions in each stage, which lead to proceed of our work.

REFERENCES


