www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 2 Issue 4 April, 2013 Page No. 1280-1285

"Digital Image Processing Approach for Fruit and Flower Leaf Identification and Recognition"

Miss. Pande Ankita V. *, Prof. Shandilya V.K.
***Department of computer science and engineering Sipna college of engineering and technology Amaravati , India**

Abstract- *Fruit and Flower trees are broadleaf and usually deciduous, meaning they lose their leaves annually in the fall. Knowing what types of fruit trees occur naturally will narrow the list of possibilities when trying to identify a tree. Observing the leaf of a tree or shrub is an excellent way to determine the species. Characteristics such as shape, size, texture and how the leaf is arranged can help identify the type of plant when fruit has not yet formed. Leaf features extraction on fruit leaf image is still be a problem on automatic plant leaf identification. This paper presents identification of local fruit trees through leaf structures using image processing techniques. In this study Chain code method is used which is used to obtain the shape of an object. Computer-Aided Plant Species Identification Technique (CAPSI) is used which is based on the image matching technique of leaf shape. The image pre-processing begins with converting the RGB image to the grey-scale image by applying thresholding technique before removing the noise. Sobel operator is applied to the binary image to recognize the edge of that image before thinning the edges. The feature extraction process is then conducted by using the chain code technique. The last stage is to recognize the leaves feature by using Linear Comparision technique.*

Keywords–Computer Aided Plant Species Identification Technique,thresholding technique,Linear-Comparision technique, Sobel operator ,. ambangan classifier.

I. INTRODUCTION

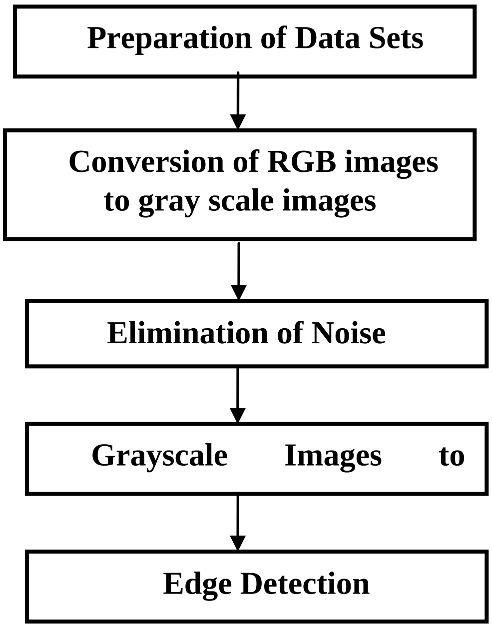
Plants play an important role in our environment. Without plants there will be no existence of the earth's ecology. But in recent days, many types of plants are at the risk of extinction. To protect plants and to catalogue various types of flora diversities, a plant database is an important step towards conservation of earth's biosphere. There are a huge number of plant species worldwide. To handle such volumes of information, development of a quick and efficient classification method has become an area of active research. In addition to the conservation aspect, recognition of plants is also necessary to utilize their medicinal properties and using them as sources of alternative energy sources like bio-fuel. There are several ways to recognize a plant, like flower, root, leaf, fruit etc. In recent times computer vision methodologies and pattern recognition techniques have been applied towards automated procedures of plant recognition. Digital image processing is the use of the algorithms and procedures for operations such as image enhancement, image compression, image analysis, mapping, geo-referencing, etc. The influence and impact of digital images on modern society is tremendous and is considered as a critical component in variety of application areas including pattern recognition, computer vision, industrial automation and healthcare industries.

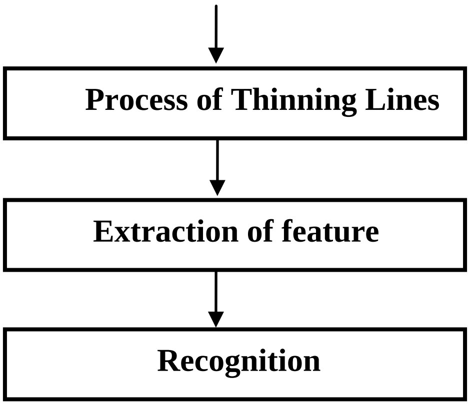
Image processing can be roughly categorized as follows: imaging, image denoising, image restoration, image coding, image segmentation, image transformation, object representation and recognition, and content-based image retrieval. Typical tasks for computer vision are scene interpretation, object recognition, optical character recognition, registration, feature extraction and video tracking. Edge detection methods utilize intensity gradients to detect the boundaries of objects. However, edge detection methods usually generate edges that are not closed contours, and this causes difficulties for later processing such as object recognition. Curve evolution methods have been popular for image segmentation since the early 1990s. These methods evolve the initialized curve(s) to the boundaries of objects in an image. The evolution of the curves may be driven by image gradient information region information or their combination .These methods are theoretically solid and numerically stable. Moreover, these methods generate image segments enclosed by closed contours, which leads to straightforward post processing. When we wander around the fields, we can find lots of plant. However, we rarely know their names.

Plants are basically classified according to shapes, colors and structures of their leaves and flowers However, if we want to recognize the plant based on 2D images, it is difficult to analyze shapes and structures of flowers since they have complex 3D structures. On the other hand, the colors of leaves are always green; moreover, shades and the variety of changes in atmosphere and season cause the color feature having low reliability. Therefore, we decided to recognize various plants by the grey-level leaf image of plant. The leaf of plant carry useful information for classification of various plants, for example, aspect ratio, shape and texture. The system is user friendly.

II. SYSTEM DESCRIPTION

The major steps are discussed in the consecutive sub-sections.





III. TECHNIQUE USED

- **Computer-Aided Plant Species Identification Technique (CAPSI)**

In this paper, 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 adopted 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. The experimental result showed that our proposed algorithm for leaf shape matching is very suitable for the recognition of not only intact but also partial, distorted and overlapped plant leaves due to its robustness.

- **Chain code technique**

A chain code is a lossless compression algorithm for monochrome images. The basic principle of chain codes is to separately encode each connected component, or "blob", in the image. For each such region, a point on the boundary is selected and its coordinates are transmitted. The encoder then moves along the boundary of the region and, at each step, transmits a symbol representing the direction of this movement. This continues until the encoder returns to the starting position, at which point the blob has been completely described, and encoding continues with the next blob in the image. This encoding method is particularly effective for images consisting of a reasonably small number of large connected components.

- **Artificial Neural Network**

Artificial Neural Network is used for the classification process. There are many studies that have been conducted previously using ANN classifier [6]. They proposed a combination of this technique and the ambangan classifier. Although ambangan is based on image histogram that could help to extract a few pixels veins, it does not show good results for different illumination conditions. Thus, the combination of ambangan methods and Artificial Neural Network (ANN) is used. This combination of features could be used to extract veins more accurately, but computational complexity is still regulated. Implementation of the proposed method is shown in Figure 4.

- **Preparation of Data Sets**

In this method images of local fruit leaves were used as the data set. Ten leaves were collected from five species of trees within the scope set. Among the tree leaves are cempedak leaves, ciku, durian, rambutan, and pulasan. Harvested leaves were kept overnight to obtain a flat leaf structure. This is so to facilitate the intake of leaf images using a digital camera. Leaves were then placed over a white cloth measuring 28 cm x 18 cm as the back-ground. The images were taken using the digital camera in the well-lit room. The distance between the camera and leaf samples was set at 23 cm. Each leaf was photographed three times to obtain more accurate information. A total of 150 leaf images were taken and used in this study. All leaves images were stored in JPEG format and will be used as the data set.

- **Conversion of RGB images to gray scale images**

In this process, RGB images are converted to grayscale image (grayscale). The formula used to convert the RGB pixel values to gray scale pixel values are as follows:

gray = 0.2989 * R + 0.5870 * G + 0 .1140 * B.

Where R is red, G is Green and B is blue represent each color pixel.

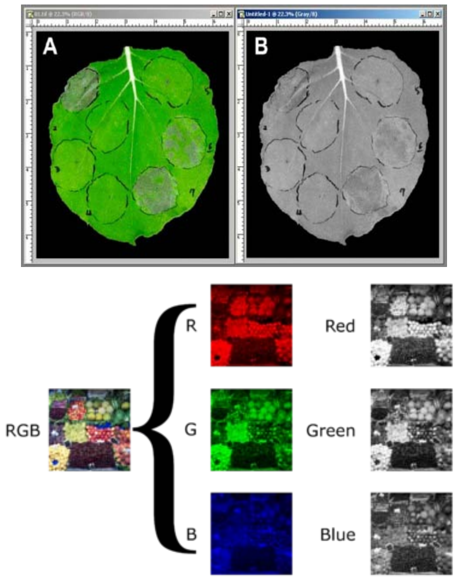


Fig 1-Sample Figure for conversion

- **Elimination of Noise (Noises Removal)**

This process is a process where the noise is removed.RGB images which were converted to gray scale image were used in this process. For noise reduction, a median filter with size 3 x 3 is used as sampler. In this process, the gray level at each pixel is replaced with the median graylevel at the neighborhood pixel. To use a median filter on the neighboring pixel, the value of these neighboring pixels and pixels arranged in ascending order.

The median of these values is obtained and set as the value in pixels. Figure 6 shows an example of the size of the 3x3 me-dian filter. The gray level values in the neighborhood are(64, 64, 64, 64, 255, 255, 64, 64, 255). These values are then arranged in ascending order (64, 64, 64, 64, 64, 64, 255,255, 255). The value 64 is arranged in the median of the gray level as highlighted in Figure 6.Fig. 6. The median filter with size 3 x 3.

- **Conversion Grayscale Images to Binary**

Taking the threshold as a level to separate the background from the leaf, the values less than threshold is taken as white and the values greater is taken as black. The leaf image is now white color and the background black, as shown in the Fig 2.

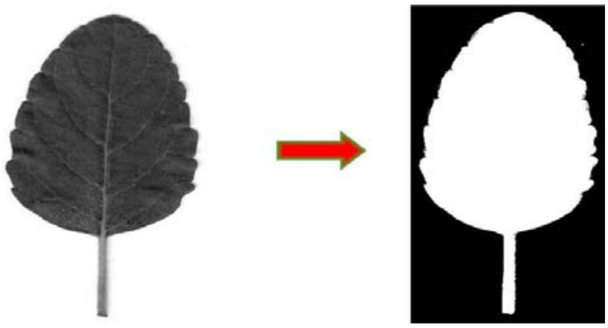


Fig 2 -Binary converted image

- **Edge Detection**

The **Sobel operator** is used in image processing, particularly within edge detection algorithms. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel operator is either the corresponding gradient vector or the norm of this vector. The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations.

On the other hand, the gradient approximation that it produces is relatively crude, in particular for high frequency variations in the image. The operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives - one for horizontal changes, and one for vertical. If we define **A** as the source image, and $\mathbf{G}_x$ and $\mathbf{G}_y$ are two images which at each point contain the horizontal and vertical derivative approximations, the computations are as follows:

$$\mathbf{G}_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} * \mathbf{A} \quad \text{and} \quad \mathbf{G}_y = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * \mathbf{A}$$

where $*$ here denotes the 2-dimensional convolution operation.

Since the Sobel kernels can be decomposed as the products of an averaging and a differentiation kernel, they compute the gradient with smoothing. For example, $\mathbf{G_x}$ can be written as

$$\begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} +1 & 0 & -1 \end{bmatrix}$$

.

- **Process of Thinning Lines**

Line thinning process is a morphological operation used to remove foreground pixels in a binary image. After the edge detection process is carried out, the resulting lines are thick which contains several pixels. Therefore ,the line thinning process is carried out to get only one line of pixels. This is required to facilitate the feature extraction process to be carried out later. If the line has a number of pixels, the feature extraction process will experience problems.

.

64	64	64	64	64	64	64	64
64	64	64	64	64	64	64	64
64	64	64	64	64	255	255	64
64	64	64	255	255	64	64	64
64	64	64	64	255	64	64	255
64	64	64	64	64	255	64	128
64	64	64	64	64	64	128	128
64	64	64	64	255	128	128	128

- **Extraction of feature**

After the pre-processing is done, feature extraction is easy. Our method takes into account only the shape of the leaf and the biometric features of the leaf.

1)Leaf width factor: of the leaf in hand is measured by slicing across the major axis and parallel to the minor axis,

2) Diameter: The diameter is defined as the longest distance between any two points on the margin of the leaf. It is denoted as D.

3) Major axis: The line segment connecting the base and the tip of the leaf is the major axis.

4) Minor axis: The maximum width, which is perpendicular to the major axis, is the minor axis of a leaf.

5) 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. For example, the area of the region in the image segment.

6) Perimeter: Perimeter of a leaf is the summation of the distances between each adjoining pair of pixels around the border of the leaf.

7) Aspect ratio: The aspect ratio is defined as the ratio of physiological length Lp to physiological width Wp, thus Lp/Wp.

8) 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.

9) 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 leaf area.

10) Narrow factor: Narrow factor is defined as the ratio of the diameter D and physiological length Lp, thus D/Lp.

11) Perimeter ratio of diameter: Ratio of perimeter to diameter, representing the ratio of leaf perimeter P and leaf diameter D, is calculated by P/D.

12) 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).

13) Contour: Also known as border following or boundary following; contour tracing is a technique that is applied to digital images in order to extract their boundary. Contour tracing is one of many preprocessing techniques performed on digital images in order to extract information about their general shape. Once the contour of a given pattern is extracted, it's different characteristics will be examined and used as features which will later on be used in pattern classification. Therefore, correct extraction of the contour will produce more accurate features which will increase the chances of correctly classifying a given pattern. First calculate the centroid .Out of the all points that represent the border32 points (p1,p2,p3,……p32) are selected. The m points are chosen so that the arcs length between any two consecutive points is equal. The Euclidean distance between the centroid c' and each of the points become the feature values.

- **Recognition**

Recognition process was carried out by comparing the length, width and perimeter of leaves that have been tested with the reference. The value of the leaf length will be matched with the long leaves that have been stored a straining data. If the length of the leaves is the same for other types of leaves, leaf width and perimeter comparison were conducted to obtain accurate results. In this study, a comparison of length, width and perimeter of the leaves is not sufficient to identify the types of leaves. Although the leaves come from different trees some of them have the same characteristics. Thus, the shape characteristic represented by the chain code numbers was used for identification purposes. Recognition by the chain code number requires other identification methods. In this study, the linear comparison method was used.

- **Linear Method Comparison**

Linear method comparison is used for leaf shapes recognition. For this process, two different sets of datasets were used which were the training data and test datasets. For the training data set, a total of seven leaves of each type of tree were used, while for the test data set ,three leaves for each type of trees were used. Comparison is the concept of Linear Dynamic Time Warping concept (DTW), which compares the distance between two time series. However, in this study, comparison of linear equations have been used and the percent-age of corresponding values between two time series were compared with the Dynamic Time Warping using time curve [10,11].The original algorithm for Dynamics Time Warping(DTW) is as follows. Based on this algorithm,

f (i, j-1) represents the insertion, the equation

f (i, j, j-1) represents the removal (deletion) the equation

f(j-1,j) represents the corresponding value (match).The corresponding value used to identify the types of features was based on the linear leaves that had been de-rived before.

IV.CONCLUSION

The recognition of local fruit trees through leaf structures using image processing techniques have been carried out. Chain code method is a method that was used to obtain the shape of an object. In this study, the chain code method has been used to extract special features of the leaves such as length, width, shape and perimeter. In addition, a linear feature recognition technique for comparison was successful implemented to achieve the objectives of the research. In this study, the researchers, through digital image processing, made a system that identifies the following: fruit tree category, provide its statistical analysis and general information using an image of a leaf as a parameter. This system gives users the ability to identify trees based on photographs of the plant's leaves taken with a digital camera. At the heart of this system is a modernized process of identification, so as to automate the way of identifying the fruit bearing trees through leaf image and digital image processing.

V. REFERENCES

[1] Abdul Kadir, Lukito Edi Nugroho, Adhi Susanto, Paulus Insap Santosa. 2011. Leaf Classification Using Shape, Color, and Texture Features, International Journal of Computer Trends and Technology

[2] X.-F. Wang, J.-X. Du, and G.-J. Zhang, "Recognition of leaf images based on shape features using a hypersphere classifier," in Proceedings of International Conference on Intelligent Computing 2005, ser. LNCS3644. Springer, 2005.

[3] D. J. Hearn, Shape analysis for the automated identication of plants from images of leaves, Taxon 58 (2009) 934{954.

[4] S. G. Wu, F. S. Bao, E. Y. Xu, Y-X Wang, Y-F Chang, and Q-L Xiang, "A leaf recognition algorithm for plant classification using probabilistic neural network", The Computing Research Repository (CoRR), vol.1, 2007, pp. 11-16.

[5] Qingfeng Wu, Changle Zhou and (2006) Chaonan Wang (2006),Feature extraction and xml representation of plant leaf for im-age Retrieval, APWeb Workshops 2006, LNCS 3842: 127-131.[10]

[6] Guang-Quan Lu, Hong-Guo Xu, Yi-bing Li (2005), Line Detection Based on Chain Code, 0-7803-9435-6/05.[9]

[7] R. Melville, The accurate de_nition of leaf shapes by rectangular coordinates, Annals of Botany 1 (1937) 673{679.