Journal of Education for Pure Science- University of Thi-Qar
Vol.10, No.2 (June, 2020)
Website: jceps.utq.edu.iq Email: jceps@eps.utq.edu.iq

DOI: http://doi.org/10.32792/utq.jceps.10.02.01

Classification of Plants by Using Computer

Walla Rahim Juadih[1] Dr. Kadhim M. Hashim[2]

[1,2]Computer Science Department, College of Education for Pure Sciences, University of Thi-Qar, Nasiriya, Iraq.

Received 17/09/2018 Accepted 06/11/2018 Published 30/11/2020

Abstract:

Computer is very important in our life and they get into all fields. So, in this research computer's application will be used and it's importantly in recognition and classification plants' leaves, because of fewness of plants' classification experts. The used data base includes (35) types of different special plants after preprocessing on these pictures, the primal processing includes two practicals': (picture enhancement and verges determine for image). In addition, we extract a group of geometrical features of plants' leaves depending on outwardness surround for leaf. Ten features are extracted. Euclidian for each leaf contains (surround, distance, center, main diameter, second diameter, and Extrema points). Traditional distance is used for measuring distance between any two points. Then, the leaf of plant is recognized using Algorithm C4.5.

Keywords: Primal processing, Features Extraction, Algorithm C4.5.

1.1 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 protection 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 protection 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, and fruit ets.

The source of database at: http\\Standard leaves.tar.gz**.**

1.2 Tree leaves:

Biologists receive a large number of requests to identify plants for people, many species of plants look very similar on their leaves, and botanists will turn to identifying the species based on their structure or other morphologies. There are three main parts to a leaf:

1. The base which is the point at which the leaf is joined to the stem.

2. The stalk or petiole is the thin section joining the base to the lamina.
3. The lamina or leaf blade is the wide part of the leaf.

1.3 Leaf shape:

There are several reasons underlying the focus on leaf shape:

1- The shape has perhaps the most discriminative power although leaves from the same plant may differ in detail, it is often the case that different species have characteristic leaf shapes, and these have often been used by botanists to identify species. Whereas differences in margin character or vein structure maybe fairly subtle, shape differences are often more obvious, even to the non-expert. In many cases, leaf size is largely determined by the environment, while shape is more heritable.

2- This is the easiest aspect to automatically extract. If a leaf is imaged against a plain black or white background, then simple threshold techniques can be used to separate the leaf from the background, and the outline can then be found by simply isolating those pixels of the leaf that border the background.

3- There are numerous shape that have already proven their worth for other biological existing morph metric techniques which can be applied to leaf problems and which may already be familiar to many botanists.

Finally, the gross structure of a leaf may be preserved even if the leaf specimen is damaged [1]. show in Figure (1)

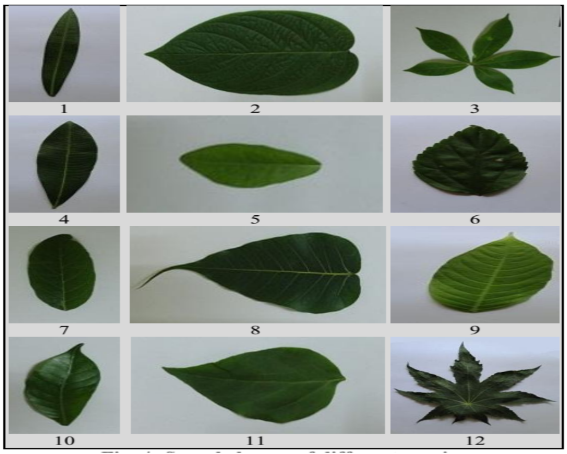

Figure (1): Example of leaf shapes

2- Related works:

In the literature, some related works were noticed so, there are some scientists' opinions such as the following:

1. In 2007, Stephen etal.[2] suggested a neural network approach for plant leaf recognition. The computer can automatically classify 32kinds of plants via the leaf images loaded from digital cameras or scanners. Brief PNN is adopted for it has fast speed on training and simple structure. 12 features are extracted and processed by brief PCA to form the input vector of PNN. Experimental result indicates that our algorithm is workable with an accuracy greater than 90% on 32 kinds of plants. Compared with other methods, this algorithm is fast in execution, efficient in recognition and easy in implementation. In this paper under consideration to improve it.

2. In 2012Piyush, etal. [3] in this research they had talked about image processing based algorithm is implemented for leaf area measurement. The system requires a digital camera, black square object and a white sheet. Images are collected in JPEG format. Images are color transformed into CIELAB color space to wipe out the noise. Threshold is calculated on „L" component of CIELAB color space by OTSU"s method. Segmented binary image is inverted in color and holes in leaf region are filled using region filling

technique. Finally leaf area is measured by pixel number statistic. The results are compared with measurements of grid count method.

3- In 2014, Arunpriya, etal. [4] Wrote about The plant types recognition is done by three phases namely the preprocessing, feature extraction and classification. At first image is preprocessed using fuzzy DE noising by Dual Tree Discrete Wavelet Transform (DT-DWT) and Boundary enhancement to remove the noise and enhance the leaf image. Then morphological and geometrical feature are extracted which is more important in clustering. Here 25 features are extracted in the feature extraction phase. These features are given as an input for the classification. The classifier algorithm is initially trained by those features. Then they are classified based on it. The ANFIS algorithm taken here is evaluated by taking 5 types of leaves and it produces 86% of accuracy in short duration. This can be further improved in future by adding features robust for leaves with irregular margins.

4. In 2016, Vandana, etal. [5] In this research suggested an approach for transforming digital morphological features to principal component analysis presented. First the images are transformed into gray scale 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.

5- In 2017 Sandesh, etal. [6] that paper provides efficient and accurate plant disease detection and classification technique by using MATLAB image processing. The proposed methodology in that paper depends on K-means and Multi SVM techniques which are configured for both leaf & fruit disease detections. The MATLAB software is ideal for digital image processing. K-means clustering and SVM algorithm provides high accuracy and consumes very less time for entire processing. In future work, we will extend our database for more plant disease identification.

---

3. Theoretical background:

There are many concepts about theories of gray scale conversion and technique of median filter. Also, they have the variety of morphology operator elements. In addition, the geometric features (Area, Distance, Perimeter, Centroid, Centroid points Extrema, Major Axis Length, Minor Axis Length). And finally, the classification the plant classification is obtained using an algorithm.

3.1 Gray scale conversion

Any image can be converted from RGB color space to a gray-scale mode using a simple transform. Gray-scale conversion is the initial step in many image analysis algorithms, as it essentially simplifies (i.e. reduces) the amount of information in the image. Although grey-scale image contains less information than a color image, the majority of important, feature related information is maintained, such as edges, regions, blobs, junctions and so on. Feature detection and processing algorithms then typically operateon the converted grayscale version of the image. The RGB color image, (Color), is converted to grey scale, using the following transformation:

$$\mathrm{Igray\_Scale}(n,m) = \alpha\,\mathrm{Icolor}(n,m,r) + \beta\,\mathrm{Icolor}\,(n.m,g) + y\,\mathrm{Icolor}(n,m,b) \quad (1)$$

Where (n, m):is indexes an individual pixel within the grey-scale image (n, m, c) the individual channel at pixel location (n, m) in the color image for channel in the red r, blue b and green g image bands the grey-scale image is essentially a weighted sum of the red, green and blue color Channels. The weighting coefficients (r, b and g) are set in proportion to the perceptual response of the human eye to each of the red, green and blue color channels and a standardized weighting ensures uniformity (NTSC television standard, α 0.2989, β 0.5780 and y 0.1140). The human eye is naturally more sensitive to red [7].

3.2 Median Filter:

Median filtering is a nonlinear smooth technology used to enhance the input image. Each pixel of gray value of neighborhood has its special pixel gray value of the median. That means all the pixels with in the neighborhood sort by gray value, taking the median of the set as neighborhood center pixel output value. Suppose the neighborhood is T, and can be expressed as follows: [8].

$$B(n, m) \ = med\ \{A(i, j), (i, j) \epsilon\ T\} \qquad (2)$$

(a) Before processing (b) after processing

Figure (2): Comparison of the image effects before and after processing using median filtering

---

3.3 Thresholding:

The simplest thresholding methods replace each pixel in an image with a black pixel if the image intensity {\display style I_i, j}} I_i, j} is less than some fixed constant T (that is, {\display style I_i, j} <T} I_i, j} <T), or a white pixel if the image intensity is greater than that constant. In the example image on the right, this results in the dark tree becoming completely black, and the white snow becoming completely white.

Thresholding is a standard method for image segmentation, which separates an object from its background by specifying a threshold value, (T). A pixel whose intensity is greater than or equal to the threshold value is claimed as an object pixel; otherwise, it becomes a background pixel. Thresholding can also be used to convert a grayscale image into a binary image.

If there is more than one object in the image, multilevel thresholding can be used. For example, if two objects exist in the image, two thresholds, T1 and T2, are utilized. The image pixel p (x, y) is classified using:

$$p(x, y) \in \begin{cases} background\ \ if\ p(x, y) < T_1 \\ objective1\ if\ T_1 \leq p(x, y) \leq T_2 \\ objective2\ otherwise \end{cases} \qquad (3)$$

The process is called global thresholding if the threshold value (or values when multilevel thresholding is used) is applied across every pixel of the image. It is called adaptive thresholding when different threshold values are applied to different areas of the image, thereby dividing the image into sub-images. Different threshold values may be used to segment each sub image. The process is called global thresholding if the threshold value (or values when multilevel thresholding is used) is applied across every pixel of the image. It is called adaptive thresholding when different threshold values are applied to different areas of the image, thereby dividing the image into sub-images. Different threshold values may be used to segment each sub image [9].

3.4 Morphological operations:

Morphological operations are widely used in computer graphics and in image analysis application. They form a well-defined system of operators that interact with shapes. Therefore, morphology has played an important part in image analysis in a variety of applications that deals with image processing **[**10**].**

The two principal morphological operations are dilation and erosion. Dilation allows objects to expand, thus potentially filling in small holes and connecting disjoint objects. Erosion shrink's objects by etching away (eroding) their boundaries.

4. Geometric features:

Geometric features are the features of the objects that have been created by a group of geometric elements like points, lines, curves or surfaces. In our proposed Algorithm then are a set of measurements (Euclidean distance, Area, Perimeter, Centroid and Extrema Points) to extract the features of the hand gesture as better than the others.

4.1 Euclidean distance:

The Euclidean distance is the space between two points in Euclidean space. The two points V and P in two-dimensional Euclidean spaces [11].

If V = (v1, v2), P = (p1, p2) are two points, then the Euclidean Distance between V and P is given by:

$$EU(V,P) = \sqrt{(v_1 - p_1)^2 + (v_2 - p_2)^2} \quad (4)$$

4.2 The Area:

The area is the extension of shapes, and it is different from the perimeter. Where the linked inside the shape, there are many known formulas for simple forms such as triangles, rectangles, and circles. Using these formulas, any polygon area can be calculated by dividing the polygon into triangles or circles to obtain curved shapes with borders and then collected after the calculation of their areas and when the polygon is irregular can polygon area is calculated by equation Gauss trapezoidal and described as in the following equation (5) [12].

$$A = \frac{1}{2}\sum_{i=0}^{n-1}(x_i * y_{i+1}) - (x_{i+1} * y_i) \quad (5)$$

4.3 The Perimeter:

The perimeter is the length of the line that surrounds the two-dimensional shapes, such as the circle, square, rectangle or irregular shapes [13].

4.4 Centroid and Extrema point:

The centroid is a fixed point in the object where the lines pass through this point which represents the weight of the object. The centroid is different from each other in terms of form or acclimatization and thus determine the status of a centroid related to this difference [14].

In mathematics and physics, the centroid or geometric center of a plane figure is the arithmetic mean position of all the points in the figure. Informally, it is the point at which a cutout of the shape could be perfectly balanced on the tip of a pin. The definition extends to any object in *n*-dimensional space: its centroid is the mean position of all the points in all of the coordinate directions While in geometry the word barycenter is a synonym for *centroid*, in astrophysics and astronomy, the barycenter is the center of mass of two or more bodies that orbit each other. In physics, the center of mass is the arithmetic mean of all points weighted by the local density or specific weight. If a physical object has uniform density, its center of mass is the same as the centroid of its shape [15].

---

5. Classification:

Classification is a method to extract the data and the division of this data into categories or specific groups in advance. It is a way to learn under the supervision of training requires disaggregated data to induce rules for the classification of categorizing data. There are two phases; the first stage is the stage of learning, which training data analysis and creating classification rules. The second stage is a classification, where the test data is classified into categories based on the rules generated [16].

5.1 C4.5 Algorithm:

C4.5 is a most important Algorithm used for decision making and creating decision trees. It is one of developments of the ID3 Algorithm used to overcome its flaws. Decision maker "decision trees" created by the C4.5 Algorithm can also be used for classification. It turns out that C4.5 is also regarded as a statistical classifier. It produces a number of change to upgrade the ID3 Algorithm. Some of these changes are mentioned.

Dealing with training data with missing values for attributes that deal with different cost characteristics in pruning decision tree after their characteristics in dealing with both kinds of values: discrete and continuous.

Consider a training data set of the already sorted samples. Each sample is Xi =S1and S2.Sn. Is a vector where S1 and S2 …Sn represent the attributes or features of the sample. Data is simply a vector D= d1 and d2. dn where d1 and d2. dn represent the class to which each sample belongs.

In each node of the tree, C4.5 chooses one attribute for data that divides the data set from more effectively into subsets that can be one class or another.

Is the natural gain of information (the difference in entropy) that results from the choice of a feature to divide the data? The attribute factor is considered with higher gain normalized information for decision making. The C4.5 Algorithm then keeps processing on the smaller sub-lists that have the maximum natural gain of information [16].

---

6. Proposed approach:

the proposed approach contains five phases, the first image acquisition, second phase is image pre-processing, the third phase segmentation, and fourth phase features extraction. The fifth phase is classification by using Algorithm (C4.5) as shown in Figure (3)

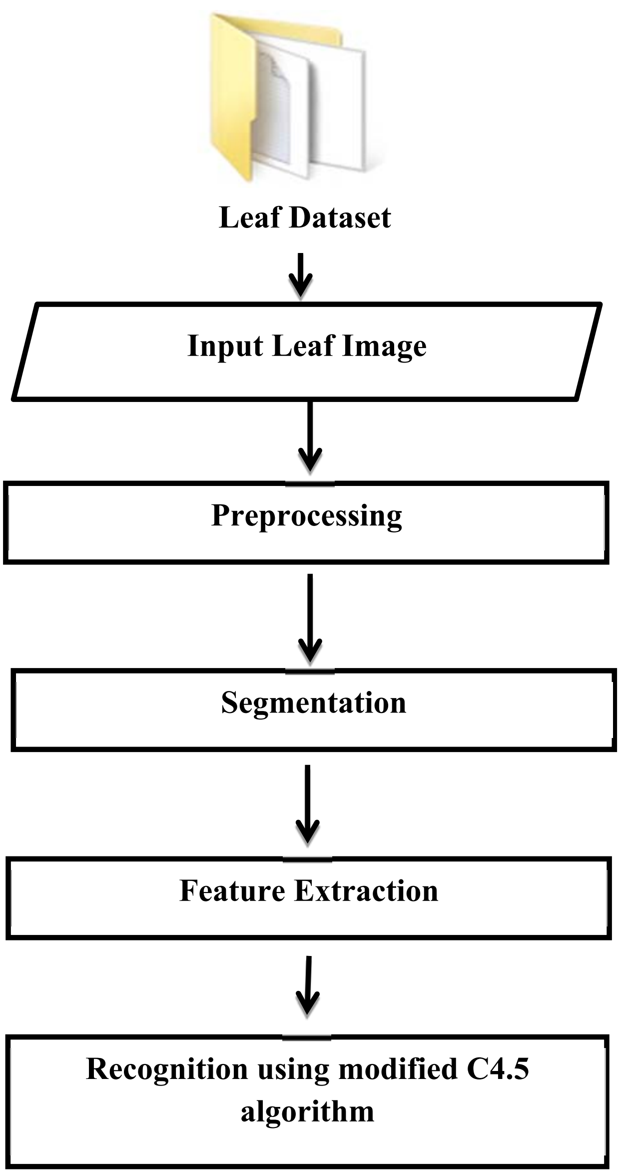

Figure (3): Main Block diagram of the proposed approach

6.1 Preprocessing:

The second main part of the proposed system is part provides an overview of the proposed algorithm for leaf differentiation in this step the image is processed and converted to gray and then get rid of noise. As shown in Figure (4)

Figure (4): Step Pre-processing Original image

Then, Convert the image after the Image median filter to the binary. As show in figure (5)

Figure (5): convert image median filter to binary image

6.2 **Morphology:**

Morphological operations are simple to use and works on the basis of set theory. The objective of using morphological operations is to remove the imperfections in the structure of image. Most of the operations used here are combination of two processes, dilation and erosion. It is applied to the binary image and the result is as shown in The Figure (6)

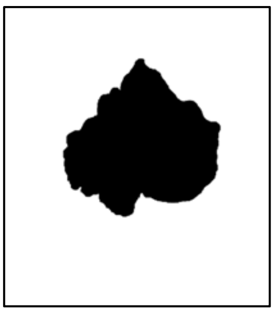

Figure (6): Morphology image

6.3 Segmentation:

Image segmentation is one of the most important processes for many applications. In image analysis and computer vision, segmenting the image correctly into real objects of interest is critical for object classification that will determine the success or failure of the algorithm. In this step, the resulting image is obtained from the morphological process and the inverse operation is carried out, the conversion of white to black and black to white. The resulting image of the reverse process is segmented of the leaf shape as show Figure (7)

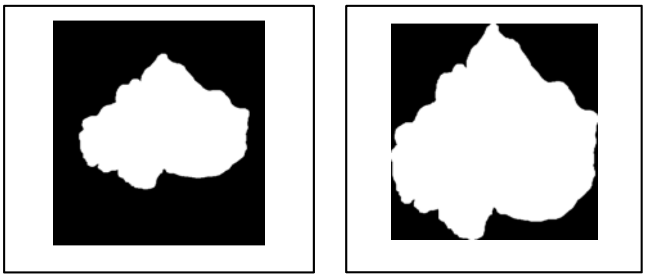

Figure (7): Segmentation image

6.4 Feature Extraction:

Step calculate Feature Extraction on Segmentation image (10- features) (8- feature calculate Euclidean distance between point center and points Extremia) , (1 feature

Figure (8): Points Extremia and point center

Plant name	D(C,E1)	D(C,E2)	D(C,E3)	D(C,E4)	D(C,E5)	D(C,E6)	D(C,E7)	D(C,E8)	Preimetre Div Area	Major Div Minor
Tilia tomentosa	1417.6093	1052.9731	147.6732	150.6955	272.0753	270.4237	151.6258	157.6088	229.0370	228.5318
Quercus Suber	228.531854	229.0371	157.6088	151.6258	270.4237	272.0753	150.6955	147.6732	0.0128	1.5955
Salix atrocinerea	296.3347	295.0303	161.7465	158.7527	313.5099	313.1205	172.4480	180.9118	0.0104	1.8653
Alnus sp	263.6694	262.3402	234.2265	227.4180	258.2531	257.0771	226.8686	228.1124	0.0089	1.0992
Quercus robur	320.1914	323.2483	180.9392	173.3460	365.0968	366.0959	162.4950	154.9902	0.0147	2.8110
Ilex aquifolium	232.7272	232.9950	116.9723	115.3579	280.7761	281.7447	115.8878	112.2010	0.0192	2.2508
Acer palmaturu	237.1398	240.6596	230.8816	225.7332	175.3756	172.0028	203.7562	202.5083	0.0244	1.0994

Table (1): Feature Extraction

At this phase, the results of the previous phase are received, which are a value of 10, divided into three groups. Group one (distance) consisting of eight values, group two (the ratio between object area to object (leaf) perimeter) consisting of one value, group three (the ratio between object Major Axis Length to object (leaf) Minor Axis Length) consisting of one value

7.CONCLUSLON:

This paper introduces modified algorithm C4-5 approach for plant and leaf recognition. The computer can classify automatically 35 kind of plants via the leaf images loaded from data base or digital cameras or scanners. C4-5 is adopted for it has fast speed on training and simple structure. Ten features are extracted and processed to form the input vectors to C4-5. Experimental result algorithm is workable with an accuracy greater than 96% on 35 kinds of plants. Compared with other methods, this algorithm is fast in execution, efficient in recognition and easy in implementation. Future work is under consideration to improve it.

Reference:

1. Bylaiah, V. (2014). *Leaf recognition and matching with Matlab* (Doctoral dissertation, Sciences) Cope, J. S., Corney, D., Clark, J. Y., Remagnino, P., & Wilkin, P. (2012). Plant species identification using digital morphometrics: A review. *Expert Systems with Applications*, *39*(8), 7562-7573.
2. Wu, S. G., Bao, F. S., Xu, E. Y., Wang, Y. X., Chang, Y. F., & Xiang, Q. L. (2007, December). A leaf recognition algorithm for plant classification using probabilistic neural network. In *2007 IEEE international symposium on signal processing and information technology* (pp. 11-16). IEEE.
3. Chaudhary, P., Godara, S., Cheeran, A. N., & Chaudhari, A. K. (2012). Fast and accurate method for leaf area measurement. *International journal of computer applications*, *49*(9), 22-25.
4. Kulkarni, A. H., Rai, H. M., Jahagirdar, K. A., & Upparamani, P. S. (2013). A leaf recognition technique for plant classification using RBPNN and Zernike moments. *International Journal of Advanced Research in Computer and Communication Engineering*, *2*(1), 984-988.
5. Patil, V. R., & Manza, R. R. (2016). Orthogonalise Digital Morphological Features Using Principal Component Analysis.
6. Khirade, S. D., & Patil, A. B. (2015, February). Plant disease detection using image processing. In *2015 International conference on computing communication control and automation* (pp. 768-771). IEEE.
7. Masoud, O. (2000). Tracking and analysis of articulated motion with an application to human motion. University of Minnesota.
8. Kumar, M., & Saxena, R. (2013). Algorithm and technique on various edge detection: A survey. *Signal & Image Processing*, *4*(3), 65.
9. Chandel, R., & Gupta, G. (2013). Image filtering algorithms and techniques: A review. *International Journal of Advanced Research in Computer Science and Software Engineering*, *3*(10).
10. Kaushik, D., Singh, U., Singhal, P., & Singh, V. (2013). Medical image segmentation using genetic algorithm. *International Journal of Computer Applications*, *81*(18), 10-15.
11. Amalorpavam, G., Naik, H. T., Kumari, J., & Suresha, M. (2013). Analysis of Digital Images Using Morphlogical Operations. *International Journal of Computer Science & Information Technology*, *5*(1), 145.
12. Stroud, A. H. (1971). Approximate calculation of multiple integrals.
13. Dr Yeap Ban Har, Dr Joseph Yeo,Teh Keng Seng , Loh Cheng Yee,Ivy Chow,Neo Chai Meng and Jacinth Liew , **“New Syllabus Mathematics Teacher's Resource Book1” ,**7th edition,OXFORD UNIVERSITY Press,2018.
14. Gupta, P. K. (2012). *Advanced dynamics of rolling elements*. Springer Science & Business Media.

15. Stroud, A. H. (1971). Approximate calculation of multiple integrals.

16-Massoud, M. A., Sabee, M., Gergais, M., & Bakhit, R. (2013). Automated new license plate recognition in Egypt. *Alexandria Engineering Journal*, *52*(3), 319-326.