ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

COMPUTERIZED EXTRACTION OF MORPHOLOGICAL AND GEOMETRICAL FEATURES FOR PLANTS WITH COMPOUND LEAVES

[1]HANEEN ALTARTOURI, [1]DUA'A ABU MAIZER, [1]ASMA IDAIS, [1]HASHEM TAMIMI, [2]RAMI ARAFEH

[1] Information Technology Department, Palestine Polytechnic University, P.O.Box 198, Hebron, Palestine.

[2]Biotechnology Research Center, Palestine Polytechnic University, P.O.Box 198, Hebron, Palestine.

E-mail: [1] htamimi@ppu.edu [2]arafeh@ppu.edu

ABSTRACT

Extracting and analyzing biological features is becoming a challenging approach for biologists and IT specialists alike. In botany, feature extraction of plant leaf using computer vision would help in classification and early diagnosis of plant diseases. The approach ranges from a simple species recognition using the features of a simple leaf to a sophisticated array of features in the case of a compound irregular leaf morphology. We propose a system for extracting features from an irregular compound-leaf, with minimum user intervention, then establish clustering into distinct similarity groups. We focused on analyzing features of a tomato compound leaf then a cluster analysis at the variety (intraspecific) taxonomic level was carried out. Experimental results of the clustering process showed that our methodology can be used to classification in the feature. From the samples that were included in the study, two major clusters; potato-type-leaves and tomato-type-leaves were revealed.

Keywords: *Feature Extraction, Image Processing, Compound Leaf, Morphological Features, Tomato Leaf.*

1. INTRODUCTION

The leaf of the plant contains plentiful amount of information for plant specialists and botanists alike. Besides that leaf shape provides essential information for plant taxonomy, it also gives a clue about the plant health and physiological status, so feature extraction of plant leaf using computer vision would help in classification and early diagnosis of plant diseases. The plant leaf is a two dimensional object that differs in blade shape between species and can discriminate between subspecies or even between different cultivars. Plant leaf can be either simple with a single leaf blade connected to a petiole or compound containing odd or even number of leaflets. An early attempt by Cholhong in [7] proposed a system for recognizing plant species from the shapes of their leaves based on contour information where plants with simple leaf were considered. A hierarchical representation of leaves was proposed. Both global, mid-most pieces and teeth of the leaves were represented in the features. The early work in [18] employed statistical discriminant analysis along with clustering using color based features as well as neural networks for recognizing pot plants. Recent developments in computer applications have lead to a new era of automated non-destructive methods in plant identification such as: [2, 3, 8, 9, 11, 13, 14, 17, 19].

Corney et al. (2012) [5] described an algorithm to automate the marginal tooth analysis of leaves from fossil and herbarium specimens in *Tilia* trees. They demonstrated an automatic feature extraction that is sufficient to identify different species of *Tilia* using a simple linear discriminant analysis, and that the features related to the leaf teeth are the most useful.

Du et al. [6] proposed efficient computer-aided plant species identification (CAPSI) approach, which was based on plant leaf images using a shape matching technique. A Douglas Peucker approximation algorithm is adopted to the original leaf shapes and a new shape representation was used to form the sequence of invariant attributes. Then a modified dynamic programming (MDP) algorithm for shape matching was proposed for the plant leaf recognition. Other interesting approaches

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to extract features from the plant leaves is the work by [1, 22] where fractal features were used. Gabor feature was used in [4, 26]. The work in [20, 10, 15] have investigated recognizing plants through geometrical features. Plants with simple leaves were addressed in these researches. A combination of texture and morphological features is used in [23] to classify plants based on leaf veins and shapes. Incorporating Zernike moments for feature descriptors was introduced in [24] and reported to enhance the accuracy of the classification. In [26], a new type of adaptive hybrid features, based on centroid-contour distance curve, is used to classify the plant leaves. Another type of derived features is in [27] where a feature point extraction method is used. Each extracted point is compared with a point in the already trained leaf image to find the class though artificial neural network.

The previous works in literature showed that there are many works on a simple leaf of different plant species; on the other hand, compound plant leaves are little addressed in literature due to their complexity in structure. A compound plant leaf represents a complex model because it differs in the number, size and arrangement of leaflets within it. An example of an irregular plant compound leaf is the leaf of tomato plant. It is odd pinnated with primary, secondary and intercalary leaflets in different sizes arranged irregularly on the leaf rachis (Figure 1).

In this study, we introduce a computerized approach that analyzes and clusters tomato plants from different cultivars based on their compound leaf morphology. The proposed methodology uses digital images of a compound leaf, then using image processing to pre-process the images, and then to extract the proposed morphological and geometrical features from a plant compound leaf image. After that the hierarchical clustering is used to do the cluster process. This paper implements the extracting features methodology with minimum user intervention, and gets benefits for the geometrical and morphological features from the previous works, such as the features have been proposed by Wu, et al. 2006 [21] and Zheng et al. [25]

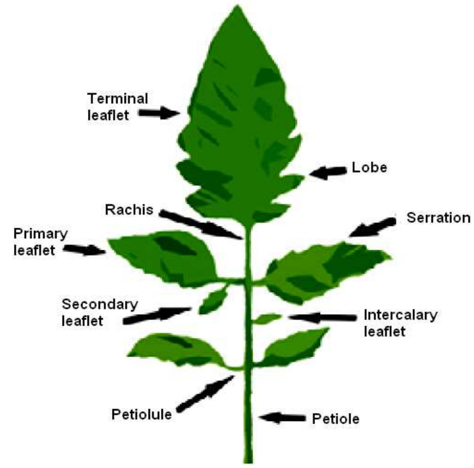

Figure 1. Diagram Of The Compound Wild-Type Tomato Leaf With Related Terminology Of The Leaf Parts. (Adopted From Berger Et Al. 2009).

2. THE PROPSED APPROACH

The literature review shows that most studies addressed plant recognition at higher taxonomic levels (intergeneric or interspecific). This study tackles a unique approach in image processing. It focuses on the clustering of plants at a deep taxonomic level (intervarietal). From the leaves included in the sample, two major clusters (potato-type-leaves) and (tomato-type-leaves) in addition to five sub-clusters in the latter could be revealed.

The accuracy of the features will be extracted depending on the image of the leaf, so in order to make the features extraction more accurate and effective some requirements s are assumed, these requirements are:

- The image should be a complete leaf of plant and not leaflet.
- The leaflets of compound leaf should not be overlaps; this can be adjusted manually by the user during the scanning process.
- The leaves should be isolated from any background.
- There is no restriction on the direction of leaves.

The proposed approach is composed of the following stages:

2.1 Image Pre-processing

Images of tomato leaves from different varieties were included in this study. The leaf image, acquired by a digital scanner or a digital camera, is binarized, filtered and scales to a given size. Since we assume there is no restriction on the direction of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

leaves, all the leaf images are automatically rotated vertically. The rotation is performed by finding the Center of Gravity (CoG), and then the Moment of Inertia (MoI) of the leaf. Given a leaf image of N pixels, we assume the function *f(m; n)*, describes the leaf surface. The CoG coordinates *(m*; n*)* can be calculated as in Equation 1 and Equations 2 [21].

$$m^* = \frac{1}{n} \sum_{(m,n)\in R}^{n} \sum m \qquad (1)$$

$$n^* = \frac{1}{n} \sum_{(m,n)\in R}^{n} \sum n \qquad (2)$$

The MoI for the image is then found as in Equation 3

$$\mu_{p,q} = \sum_{I} \sum_{J} (i - m^*)^p (j - n^*)^q \qquad (3)$$

where $p,q = 0,1,2$.

The angle θ is then calculated as in Equation 4:

$$\theta = \frac{1}{2} tan^{-1} \left[\frac{2.\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right] \qquad (4)$$

Finally, the rachis is removed from the leaf using a sequence of binary image-based morphological operations. The image pre-processing is illustrated in Figure. 2.

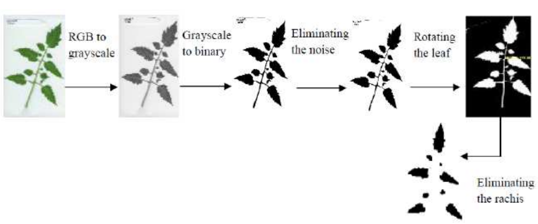

Figure 2. An Image Of A Tomato Compound Leaf In The Pre-Processing Stages.

2.2 FEATURES EXTRACTION

Several morphological and geometrical features are extracted from the compound leaf. These features provide significant information about its visual representation. Two types of features are extracted: global features, which are extracted from the compound leaf as a whole, and local features, which are extracted from each leaflet. The following subsections explain the process of extracting each type.

2.2.1 Global features. Global features have two components: the number of leaflets and the general structure of the leaf. When counting the number of leaflets, we assume there are large leaflets and small leaflets within the same leaf. Courting each of them separately produces more distinctive features than courting both of them together. In order to find the leaflets in an image, the connected component technique is applied. The connected components technique provides each contagious set of pixels, i.e. leaflet, in the image with a certain label.

Once the connected components are labeled, we calculate the area of the component. After that, we classified the leaflet as either a small leaflet as follows. Given a connected component c, the function class (C) Є {small; large} is defined as follows:

$$class(c) = \begin{cases} large, & \text{if } Area(c) \le \tau \\ small, & \text{if } Area(c) > \tau \end{cases} \qquad (5)$$

where τ is defined experimentally. Finally small leaflets and large leaflets are counted separately. The general structure of the plant leaf can be extracted by firstly detecting the center of each leaflet, then establishing a graph representation from the leaf. Each node in the graph represents the distance between the COG of the leaf (see Equations 1, 2) as a whole and the centers of each leaflet as illustrated in Figure 3.

2.2.2 Local features. The local features are extracted from each leaflet. We establish a vector of features for each node as in Figure 4. The vector contains the following features: Leaflet area (A): leaflet area is calculated by counting the number of pixels with a foreground color in the smoothed leaf image.

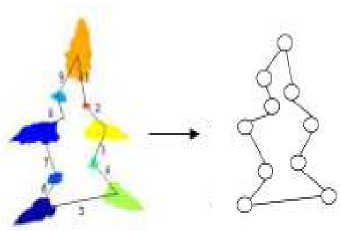

Figure 3. Extracting The Structure Of The Leaf By Establishing A Graph Of Leaflets

Journal of Theoretical and Applied Information Technology
30th November 2015. Vol.81. No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

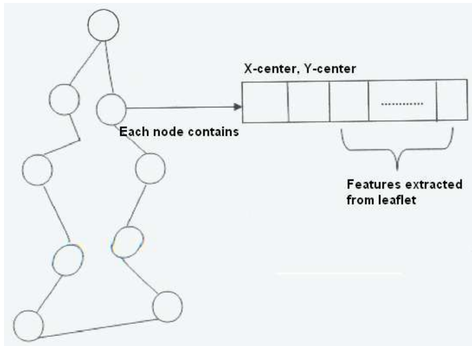

Figure 4. Graph For The Essential Features Of A Compound Leaf

—Leaflet perimeter (P): leaflet perimeter is calculated by counting the number of pixels consisting leaf margin.

—Aspect ratio (S): the aspect ratio is defined as the ratio of the physical length (Lp) to the physical width (Wp), which are calculated depending on the axes ellipse contained the leaflets, it is defined as in Equation 6:

$$S = \frac{L_p}{W_p} \qquad (6)$$

—Form factor (F): the form factor is used to describe how the shape of the leaflet is different from a circle, it is defined as in Equation 7.

$$F = \frac{4\pi A}{P^2} \qquad (7)$$

—Rectangularity (R): rectangularity is used to describes how the shape of the leaflet is different from a rectangle, it is defined as in Equation 8.

$$R = \frac{L_p W_p}{A} \qquad (8)$$

Angle between leaflet and the vertical axis (θ): Although the plant is previously rotated vertically, as discussed in Section 2.1, the deviation from the vertical axis plays an important role as a feature. The deviation changes according to the straightness of the plant rachis. This is calculated using the same method as in Equation 4.

2.3 Hierarchical Clustering of Plant Compound Leaf: clustering plant into defined groups is an important aspect in the plant taxonomy; it is used to reduce the large number of plants types of specific plant compound leaf to help in classification, and it can be used to find the similarity in the groups of plants that have compound leaves [16]. In this paper hierarchical clustering is used to achieve the clustering process because it may correspond to meaningful taxonomy, and does not have to assume any particular number of clusters, any desired number of clusters can be obtained by cutting the dendrogram at the proper level [16].

In this paper an agglomerative approach of hierarchical clustering is used to cluster the plant leaves. The Hierarchical Clustering [16] is applied using the following steps: First, the Euclidean distances between each pairs of features *ai* and *bi* is computed, as in Equation 9.

$$d_{a,b} = \sqrt{\sum_i (a_i - b_i)^2} \qquad (9)$$

After that, a tree from the calculated distances using average linkage is created, and the output is drawn as a dendrogram. But Before doing cluster process, the matrix of plant's features should be prepared. This matrix contains number of rows equals the number of leaves in the sample, each row contains the main features of the leaf (global and local features).

In this study this phase is considered difficult because each leaf contains number of leaflets and each leaflet has different number of features as well. So the dimensionality of matrix will vary and this is a main problem.

All these reasons make the process of clustering the compound leaves more difficult than a simple leaf. To overcome these problems, the distance matrix is prepared as follows:

(1) Each leaflet belongs to leaf in the sample is compared with other leaflets in other leaves using Euclidean distance of leaflet's features.

(2) Then the minimum distance between each leaflet in the leaf and other leaves is taken and stored.

(3) The minimum of distances between each leaflet in the leaf and other leaves are prepared for each leaflet.

(4) Finally, the minimum distances between each leaf and others are stored.

These steps enabled us to find the matrix *M* of leaves features; *M* is a squared matrix define as follows:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$M(i,j) = \begin{cases} 0 & if \;\; i = j \\ \delta_{i,j} & if \;\; i \neq j \end{cases} \quad (10)$$

Where $\delta i,j$ is the minimum distance between the feature of leaf i and the feature of leaf j.

3. EXPERIMENTS AND RESULTS

(1) Extracting Features Experiments and Results

We applied the features extraction procedure on 29 different tomato leaves from different varieties. High resolution images of the leaves are provided in a supplementary file and can be provided upon request.

(2) Hierarchical clustering Experiments and Results
We have applied Hierarchical clustering experiments on a sample of 29 tomato leaves from different types in this paper by using the statistical toolbox in MATLAB [14]. The output shows the dendrogram with two main clusters colored with red and blue (Figure 5).

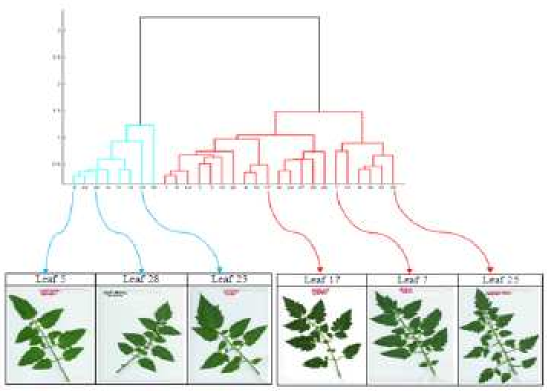

Figure 5. A dendogram of the hierarchical clustering results in two main clusters. A sample of three random leaves from each cluster is shown.

Figure 5 presents a dendogram after applying hierarchical clustering to the samples, this figure shows a clear potato-leaf tomato-leaf distinct clustering (colored with red and blue). It can be seen that the leaves in the blue cluster belongs to Potato Leaflet (PL) type that contains smooth edges and the leaves in the red cluster belongs to Regular Leaflet (RL) type that contains a toothed and symmetrical edges.

The result of clustering confirms that the features that have been extracted in this paper play an important role in the representation of the leaf's shape in accurate way, and this result will help in the classification of the plant compound leaves.

4. CONCLUSION

This paper introduces a new methodology to extract morphological and geometrical features from a plant with compound leaf. Hierarchical Clustering of plant compound leaf is applied on the sample to reduce the large number of plants types of specific plant compound leaf to help in classification, and it can be used to find the similarity in the groups of plants that have compound leaves. Experimental results of the clustering process shows that our methodology can be used to classification in the feature.

Unlike the previous studies which addressed plant recognition at higher taxonomic levels, this study tackles a focused on the clustering of plants at a deep taxonomic level (inter-varietal). From the leaves included in the sample, two major clusters (potato-type-leaves) and (tomato-type-leaves) in addition to five sub-clusters in the latter could be revealed

In the future, many changes could help in obtaining better results; by extracting new set of features such as, smooth factor, narrow factor, vein features,…etc. to represent the compound leaf accurately, and that reflect other characteristics of the plant. Also, theses extracted features can be used in the classification of plants that have compound leaves that represent an important component in human life, by using machine learning.

5. REFERENCES

[1] J. J. Camarero, S. Siso, and E. G. Pelegrin. "Fractal Dimension Does not Adequately Describe the Complexity of Leaf Margin in Seedlings of Quercus Species", *Anales del Jardin Botanico de Madrid*, Vol. 60, No. 1, 2003, pp. 63–71.

[2] J. Chaki and R. Parekh "Plant Leaf Recognition Using Shape Based Features And Neural Network Classifiers", International Journal of Advanced Computer Science and Applications", Vol. 2, No. 10. 2011, pp.: 41–47.

[3] J. Chaki and R Parekh "Designing An Automated System for Plant Leaf Recognition". International. Journal of Advances in Engineering and Technology Vol. 2, No. 1, 2012, pp:149–158.

[4] J. Cope, P. Remagnino, S. Barman, and P. Wilkin, "Plant Texture Classification Using Gabor Co-occurrences", In: G. Bebis, R.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Boyle, B. Parvin, D. Koracin, R. Chung, R. Hammound, M. Hussain,, T. Kar-Han, R. Crawfis, D. Thalmann, D. Kao, and Lisa Avila, eds, Advances in Visual Computing, Vol. 6454 of Lecture Notes in Computer Science, 2010, pp. 669–677. Springer Berlin Heidelberg.

[5] D.P.A. Corney, J. Y. Clark, H. T. Tang, and P. Wilkin. "Automatic Extraction of Leaf Characters from Herbarium Specimens". Taxon, Vol. 61, No. 1, 2012, pp. 231–244.

[6] J. X. Du, D. S. Huang, X. F. Wang, and X. Gu. "Computer-aided Plant Species Identification (CAPSI) Based on Leaf Shape Matching Technique". Transactions Institute of Measurement And Control, Vol. 28, No. 3, 2006. pp. 275–284.

[7] C. Im, H. Nishida, and T. L. Kunii, "A Hierarchical Method of Recognizing Plant Species by Leaf Shapes", In IAPR Workshop on Machine Vision Applications Makuhari, Chipa, Japan. 1998, Pp. 158–161.

[8] M. F. Ab Jabal, S. Hamid, S. Shuib, and I. Ahmad. "Leaf Features Extraction and Recognition Approaches to Classify Plant", Journal of Computer Science, Vol. 9, No. 10, 2013, pp. 1295–1304.

[9] L. Zhang, J. Kong, X. Zeng, J. Ren, "Plant Species Identification Based on Neural Network", Natural Computation, Vol. 5. 2008. pp. 90-94.

[10] N. MacLeod. "Shape Models As A Basis for Morphological Analysis in Paleobiological Systematics: Dicotyledonous Leaf Physiography", Bulletins of American Paleontology, Vol. 369. 2005. pp. 219-238.

[11] M. Shabanzade, M. Zahedi, and S. Amin Aghvami. "Combination of Local Descriptors and Global Features for Leaf Recognition", Signal and Image Processing: An International Journal, Vol. 2, No. 3. 2011, pp. 23–31.

[12] MATLAB. version 7.10.0 (R2010a). The Math Works Inc. Natick, Massachusetts, 2010.

[13] J.D. Mauseth. "Botany: An Introduction to Plant Biology", Botany and Plant Biotechnology Series. Jones and Bartlett Publishers, 2009.

[14] O. Mzoughi, I. Yahiaoui, and N. Boujemaa. "Extraction of Leaf Parts by Image Analysis". In Proceedings of the 9th international conference on Image Analysis and Recognition –Vol. 1, 2012, pp. 348–358. Berlin, Heidelberg, Springer-Verlag.

[15] T. Panagiotis, P. Stelios, and M. Dimitris. "Plant Leaves Classification Based on Morphological Features and A Fuzzy Surface Selection Technique". In: Fifth International Conference on Technology and Automation, Thessaloniki, Greece, 2005, pp. 365–370.

[16] P. N. Tan, M. Stinbach, and V. Kumar,. "Introduction to Data Mining" Addison Wesley Companion, 2006.

[17] Y. Qing, L.Dong, Z. Dongyan, and W. Xiu, "Recognition Algorithm for Plant Leaves Based on Adaptive Supervised Locally Linear Embedding", International Journal of Agricultural and Biological Engineering, Vol. 6, No. 3, 2013, pp. 52-57.

[18] A. J. M. Tmmermans and .A. A. Hulzebosch. "Computer Vision System for on-line Sorting of Pot Plants Using An Artificial Neural Network Classifier", Computers and Electronics in Agriculture, Vol. 15, No. 1. 1996, pp. 41–55.

[19] V. Satti,A. Satya, and S. Sharma, "An Automatic Leaf Recognition System for Plant Identification Using Machine Vision Technology". International Journal of Engineering Science and Technology. Vol. 5, No. 4, 2013, pp. 874-879.

[20] V. Visosi, and A. Cardini, "Leaf Morphology, Taxonomy and Geometric Morpho-Metrics: A Simplified Protocol for Beginners". PLoS ONE, Vol. 6, No. 10, 2011, e2563010.

[21] Q. Wu, C. Zhou, and C. Wang, "Feature Extraction and Automatic Recognition of Plant Leaf Using Artificial Neural Network" Advances in Artificial Intelligence - Research in Computing Science, Vol. 20, 2006, pp. 3-10.

[22] J. X.Du, C. M. Zhai, and Q. P. Wang, "Recognition of Plant Leaf Image Based on Fractal Dimension Features", Neurocomputing, Vol. 116, 2013, pp 150 – 156.

[23] Kue-Bum Lee and Kwang-Seok Hong, "An Implementation of Leaf Recognition System using Leaf Vein and Shape" International Journal of Bio-Science and Bio-Technology Vol. 5, No. 2, April, 2013

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[24] Pallavi P1, V.S Veena Devi2, "Leaf Recognition Based on Feature Extraction and Zernike Moments" International Conference on Advances in Computer and Communication Engineering, 21-22 April 2014.

[25] Xiaodong Zheng and Xiaojie Wang, "Leaf Vein Extraction Based on Gray-scale Morphology", I.J. Image, Graphics and Signal Processing, 2, 25-31, 2010.

[26] Lihua Fu, Meng Zhang, Zhihui Liu, Hongwei Li, Tingting He, "Adaptive Hybrid Feature Extraction for Leaf Image Classification by Support Vector Machine", Journal of Convergence Information Technology(JCIT) Vol7, No 2, Feb. 2012.

[27] S. R. Deokar, P. H. Zope and S. R. Suralkar "Leaf Recognition Using Feature Point Extraction and Artificial Neural Network", International Journal of Engineering Research & Technology, Vol.2 - Issue 1, Jan. 2013.

[28] Dominik L. Michels, Sven A. Giesselbach, Thomas Werner, Volker Steinhage "On Feature Extraction for Fingerprinting Grapevine Leaves" The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing, July 23, 2013.