

Leaf Classification Using Convexity Measure of Polygons

Jules Raymond Kala^(✉), Serestina Viriri, Deshendran Moodley,
and Jules Raymond Tapamo

University of Kwazulu-Natal, Durban, South Africa
raymondkala1@gmail.com
<http://www.ukzn.ac.za>

Abstract. Plant taxonomy is a long-standing practice in botany. It uses the morphology of plant leaves to make categories. Leaf shape is one of the physical characteristics used to discriminate between plant species. This paper presents the characterisation of a leaf shape using the Convexity Measure of Polygons and the seven invariant moments in combination with other morphological features to improve leaf classification. The Convexity Measure of Polygons used in this paper is based on the minimum ratio obtained by dividing the rotated leaf-bounding perimeter of the associate bounding rectangle of the leaf shape. The proposed model is rotation, translation and scale invariant. It achieves a classification rate of 92 % on 400 leaves of 20 species, 99 % on 100 leaves of 4 species and 95 % on 1600 leaves of 32 species using a Multilayer Perceptron classifier. The proposed method out-performs several state-of-the-art methods when tested under the similar conditions, even with deformed leaves.

Keywords: Shape · Minimum Bounding Rectangle (MBR) · Feature extraction · Convexity measure

1 Introduction

Life on earth depends on plants and the life cycle of human beings is dependent on them as they have the ability to transform light into food [14]. Plants were initially organised as useful and not useful. This can be considered as the first method of plant classification. The science of botany was then created and its main objectives were to create rules for the classification of plants.

In the 18th century, Linnaeus [11] developed the systematic classification of plants based on plant morphology. Given the number of rules used in the systematic classification of plants, it is a difficult task even for a trained botanist, because it is possible to find two different plants species with almost the same physical appearance.

The difficulties encountered in manual classification have prompted the necessity of computer vision to automate the process. The classification of plants is important for botanists and environmentalists who are interested in obtaining an

organised series of descriptors or features that describe plant structures. These can be used to determine the properties of plants, prevent their extinction and the spread of diseases caused by plant pollen [6]. Leaf features include texture, vein, shape and water retention capacity. These are important components to take into consideration during leaf studies.

This research proposes a leaf feature extraction model that combines the Convexity Measure with geometrical and morphological features to improve the classification rate of plant leaves.

The Convexity Measure of Polygons is a shape characteriser used to describe the overall structure of a given shape. The New Convexity Measure of Polygons was created to provide solutions to the problems observed when using the Convex Hull polygon when evaluating the convexity measure of a given shape. These problems were the detection of small variations on the shape and the calculation of the convexity measurements of shapes with holes.

The rest of the paper is organised as follows: Sect. 2 discusses some related works; Sect. 3 presents the methods and techniques used; experimental results and discussion are discussed in Sect. 4 and the conclusion and future work are presented in Sect. 5.

2 Related Work

Many approaches have been used to perform plant classification. The following sections explore and discuss the different approaches used in the literature.

2.1 Morphological Based Approach

Panagiotis et al. [12] proposed one of the morphological approaches used for plant classification based on leaf analysis. The goal of this approach was the design of a system that is able to extract specific morphological and geometrical features from a plant leaf. They further use fuzzy surface selection to select the relevant features. This approach reduces the dimensionality of the feature space leading to a very simplified model that is more adapted to real time classification application. The model obtained is scale and orientation invariant and yields to a system that achieved a classification rate of 99%, even with deformed leaves. The main drawbacks of this approach are the size (less than 5 species) and the system is not translation invariant.

Stephen et al. [13] presented a method based on the combination of the Probabilistic Neural Network (PNN) and image processing techniques to construct general purpose semi-automatic leaf recognition for accurate plant classification. This method achieved 90% accuracy, and is good in running time. However, it is not rotation invariant and requires human intervention during the process of feature extraction.

2.2 Texture Based Approaches

Esma et al. [8] proposed a system based on Dendritic Cell Algorithms derived from the Danger Theory [4], as a classifier. The wavelet transformed is used to extract the leaf features for the classification algorithm. This approach achieved a classification accuracy of 94 %.

Ahmed et al. [3] developed an approach that combines texture features based on Discrete wavelet transformation with an entropy measurement to construct an efficient leaf identifier. An accuracy of 92 % was achieved. The main advantage of this approach is noise removal from the image background. The drawback of the method is the size of the data set as it is based on less than 10 species.

2.3 Hybrid Approaches

Anant et al. [5], presented an approach based on the shape features combined with texture features to generate a feature set that will be used by the nearest neighbour classifier for the classification process. The method achieved a classification rate of 91.5 % for 14 plant species.

3 Materials and Methods

In many applications it is always important to enhance the quality of the captured image before any real processing can start. Particularly, in leaf classification the process we used will be run using four steps - input, preprocessing, feature extraction and classification.

3.1 Image Preprocessing and Leaves Data Set

Figure 1 presents samples of the categories of leaves randomly selected for the experimentation. Figure 2 shows the transformations performed on leaf image before extracting features.

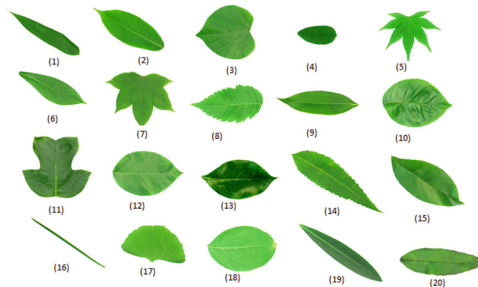


Fig. 1. Leaves selected for the classification

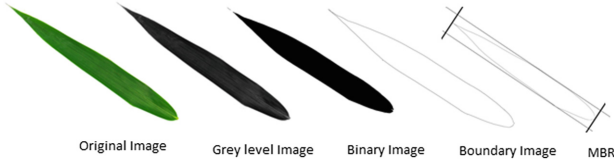


Fig. 2. Leaf image processing (Minimum Bounding Rectangle (MBR))

Image Preprocessing

- All images are transformed from a colour image into a grey level image using Eq. (1). In fact, converting the image into grey level will preserve the shape of the leaf; thereby not impacting negatively on the end result.

$$I = 0.2989 * R + 0.5870 * G + 0.1140 * B \quad (1)$$

- Leaf boundary is extracted by applying the Sobel filter on the leaf binary image.
- Thinning operation is then used to have leaf contours that are one pixel thick.

3.2 Features Extraction

The Convexity Measure of Polygons: A set of points A is said to be convex if the straight line segment joining any two points in A is contained in A [10]. The Convexity Measure of Polygons is a numerical value used to represent the probability that a straight line joining two points in A lies entirely in A .

The Convexity Measure of Polygons has the following properties as defined in [10]:

- The value of the convexity measure is in $(0,1]$.
- For a given shape, the convexity measure can be arbitrarily close to 0.
- The convexity measure of a convex set is equal to 1.
- The convexity measure is invariant under similarity transformation.

In the literature there are two types of convexity measure: surface base convexity measures and boundary base convexity measure [10]. The first approaches for the determination of the convexity measure of polygon was based on the convex Hull polygon (CH). C_1 , C_2 and C_3 were defined as:

$$C_1 = \frac{Area(S)}{Area(CH(S))}. \quad (2)$$

C_1 is a surface based convexity measure, obtained by dividing the area of the shape with the surface of the associate convex hull polygon.

$$C_2 = \frac{Area(MCS(S))}{Area(S)}. \quad (3)$$

C_2 is a surface based convexity measure, obtained by dividing the area of the minimum convex set (MCS) of shape S with the surface of shape S .

$$C_3 = \frac{Per(CH(S))}{Per(S)}. \tag{4}$$

C_3 is a boundary based convexity measure, obtained by dividing the perimeter of the convex hull of shape S with the perimeter of shape S .

The Convexity Measure of Polygons New Definition: The new definition of the Convexity Measure of Polygons introduced by J. Zunic et al. [10] was designed because of the incapacity of other convexity measure’s to include huge defects. In addition, it can evaluate small variations on a shape. It is the first element in the leaf feature vector. It is the only convexity measure used in this paper. The Convexity Measure of Polygons defined by J. Zunic et al. [10] is evaluated as:

$$C(P) = \min_{\alpha \in [0, 2\pi]} \frac{Per_2(R(P, \alpha))}{Per_1(P, \alpha)}, \tag{5}$$

where:

- α = Rotation angle
- P = Shape Parameter
- R = The optimal rectangle
- Per_2 = Perimeter by projection on axis
- Per_1 = Euclidian perimeter

In this equation the perimeter of the polygon P is fixed and the perimeter of the bounding rectangle note $R(P, \alpha)$ depends on the value of α . $C(P)$ is equivalent to the following equation.

$$C(P) = \min \left\{ \frac{Per_2(R(P, \alpha_i))}{Per_1(P, \alpha_i)} \mid i = 1, 2, \dots, n \right\}$$

Where

$$\begin{aligned} Per_2(R(P, \alpha_i)) &= g_i * \cos(\alpha_i) + f_i * \sin(\alpha_i), \\ Per_1(P, \alpha_i) &= c_j * \cos(\alpha_i) + d_j * \sin(\alpha_i). \end{aligned}$$

g_i, f_i, c_j, d_j are the constants associated to the Euclidian length of the rectangle edge and the polygon edges.

This equation represent the computation process of $C(P)$.

The Convexity Measure of Polygons (J. Zunic et al. [10]) in order to be used for shape characterisation has to be combined to other convexity measure to increase the recognition rate [10]. The Convexity Measure of Polygons only expresses how convex or concave a given shape is, but it is also important to have another descriptor for the surface.

The Seven Invariant Moments. Hu's seven invariant moments are computed from the central moments. They are very useful for shape description and classification [9]. The discrete form of the geometrical moment of order $p+q$ is defined as:

$$M_{pq} = \sum_{x=1}^N \sum_{y=1}^M x^p y^q. \quad (6)$$

where:

$$p, q = 0, 1, 2, \dots$$

$N \times M$ = the image size.

Consequently, a set of seven invariant moments (Ph_1, Ph_2, \dots, Ph_7) can be derived from the normalized central moments as in [13].

Geometrical Features. The *rectangularity* (R) represents the ratio between the leaf area (A_{leaf}) and the area of the minimum bounding rectangle. It evaluate how the leaf shape is close to a rectangle shape.

$$R = \frac{A_{leaf}}{D_{max} \times D_{min}} \quad (7)$$

The *aspect ratio* (A) is the ratio between the maximum length (D_{max}) and the minimum length (D_{min}) of the minimum bounding rectangle

$$A = \frac{D_{max}}{D_{min}} \quad (8)$$

The *sphericity* (S) is express by the following equation.

$$S = \frac{r_i}{r_c} \quad (9)$$

where:

r_i = represents the radius of the in cycle of the leaf.

r_c = the radius of the ex-circle of the leaf.

The ratio between the length of the main inertia axis and the minor inertia axis of the leaf, determines the accent of the leaf. It evaluates how much an iconic section deviates from being circular [2].

$$E = \frac{E_A}{E_B} \quad (10)$$

The *circularity* (C) is defined by all the contour points of the leaf image.

$$C = \frac{\mu_R}{\sigma_R} \quad (11)$$

where

$$\mu_R = \frac{1}{N} \sum_{i=0}^{N-1} \|(x_i, y_i) - (\bar{x}, \bar{y})\|$$

and

$$\sigma_R = \frac{1}{N} \sum_{i=0}^{N-1} (|| (x_i, y_i) - (\bar{x}, \bar{y}) || - \mu_R)^2$$

Form Factor (F) compares the perimeter of the equivalent circle to the perimeter of the leaf shape. It is also used to describe surface irregularity. It is given by the following equation:

$$F = \frac{4\pi A_{leaf}}{P_{leaf}^2} \quad (12)$$

Area ratio of convex hull (CA) is defined as the ratio between the leaf area and the area of it's associated convex Hull polygon (equivalent to the surface based convexity measure). It is expressed by the following equation:

$$CA = \frac{A_C}{A_{ROI}} \quad (13)$$

4 Experimental Result and Discussion

4.1 Experimental Results

The experiments were conducted using FLAVIA. The leaf database is composed of more than 2500 plants leaves of more than 30 species [13]. We randomly chose 400 leaves of 20 species, 1600 leaves of 32 species and 100 leaves of 5 species (the 1600 leaves of 32 species represent all the available species in FLAVIA). The experiments are organized in two phases. First the leaves are characterized using the geometrical features and the seven invariants moments. Secondly the seven invariants moments and the geometrical features are combined to the New Convexity measure of polygones to characterized a leaf image.

Table 1 presents the accuracy of the proposed method with and without the Convexity Measure of Polygon (J. Zunic et al.) in the feature vector. In the first row with the Convexity Measure of Polygons, the Multi Layer Perceptron (MLP) achieved an average of 92% of well classified leaves with an area under Received Operating Characteristics curve (ROC) equaled 0.993. Without the Convexity Measurement of Polygons, 86% of well classified leaves with the area under the ROC curve equaled 0.98.

4.2 Discussion

We decided to organise our experiment using three processes to show how efficient our model is when used in various conditions. As Panagiotis et al. in [12] we used 100 images of leaves to illustrate how the proposed method is more efficient when applied to a small number of leaves. We obtained a classification rate of 99% with the Convexity Measure of Polygons (J. Zunic et al.) in the feature set it shows that the Convexity Measure of Polygon contributed significantly in the discrimination process of leaf shapes even with a small dataset.

Table 1. Comparative study of classifiers

	With the convexity measure		Without the convexity measure	
	% of good classify	AUC average	% of good classify	AUC average
Multilayer Perceptron (400)	92 %	0.993	86 %	0.98
K-Nearest Neighbour (400)	87.5 %	0.971	85 %	0.921
Naive Bayes (400)	80.94 %	0.969	79.2 %	0.966
Multilayer Perceptron (100)	99 %	0.993	97 %	0.97
K-Nearest Neighbour (100)	99 %	0.971	97 %	0.9
Naive Bayes (100)	90 %	0.969	80.2 %	0.95
Multilayer Perceptron (1600)	95 %	0.993	92 %	0.99
K-Nearest Neighbour (1600)	92 %	0.971	89.5 %	0.95
Naive Bayes (1600)	89 %	0.969	80.2 %	0.98

We then applied the proposed method to a medium dataset which had 400 leaves. In this case, a classification rate of 92 % shows again how efficient the proposed method is when applied to a medium size dataset when the Convexity Measure of Polygons (J. Zunic et al.) is used. Finally, to complete our experiment we applied the proposed to a large dataset with 1600 leaves and the classification rate obtained was 96 %. This classification rate shows that the proposed method remains consistent even with a larger dataset. The classification rate and the Area Under the ROC Curve (AUC) clearly shows that the Convexity Measure of Polygons (J. Zunic et al.) contributes to the improvement of the classification rate and to the efficiency of the proposed model.

5 Comparative Study of the Proposed Method with the Available Method in the Literature

In Table 1 some methods of plant classification using leaves are presented. In Table 2 in the first row is - the authors name; the title of the article; the nature of the features; the number of leaves used for the experimentation; the number of species; the classification rate; the classification algorithm; drawbacks of the method and advantages. On the second row of Table 2, Jixian et al. [9] proposed leaf shape based plant species recognition. This approach is shape based as 400 leaves of 20 species of plant were used and a classification rate of 91 % was achieved with the MMC. This was after reduction of the feature space in order to obtain a fast classification method. In the last row the method use in this paper is present. This method is based on the introduction of the Convexity Measure of Polygons to boost the leaf recognition rate. 400 leaves of 20 different species are used for the experiment, and the MLP achieved a classification rate of 92 %. The drawback of this method is time taken to calculate the convexity measure (using J. Zunic et al. [10] definition) of a big leaf image shapes. Despite this, the method is very accurate and fast during the classification and achieved better results even with deformed leaves.

Table 2. Table of some of the methods of plants classification using leaves

Autors	Method	Feature	Nb Leaves	Nb species	Rate	Algorithm	Drawback	Advantages in classification
Panagiotis et al [14]	Plants leaves classification based on morphological feature	Shape	100	4	99%	NN Fuzzy	incomplete experiment	Fast classification
Jixian et al [10]	Leaf shape based plant species recognition	Shape	400	20	91%	MMC	Model evaluation	Fast with the MMC
Stephen et al [15]	A leaf recognition Algorithm for plant classification using PNN	Shape	1800	32	90%	PNN	the method is not completely automatic	Fast classification
Benoit et al [7]	Weed leaf recognition in complex natural scene by model guided edge pairing	Shape	10	1	60%	rating process	incomplete methode	Adaptative method
A. H. Kulkarni al [1]	A Leaf Recognition System for Classifying Plants Using RBPNN and pseudo Zernike Moments	Shape, colour, texture	vein, 1600	32	94.52%	RBPNN	-	Fast
Proposed approach	Leaf Classification using Convexity Measure of Polygons	Shape	400	20	92%	MLP	Convexity measure time complexity with big images	Accurate
Proposed approach	Leaf Classification using Convexity Measure of Polygons	Shape	1600	32	95%	MLP	Convexity measure time complexity with big images	Accurate
Proposed approach	Leaf Classification using Convexity Measure of Polygons	Shape	100	4	99%	MLP	Convexity measure time complexity with big images	Accurate

6 Conclusion

This paper presented the classification of leaf images using shape analysis. Here the investigation of the use of the Convexity Measure of Polygons in the process of leaf shape characterisation is performed. The result obtained is a method with the following properties: rotation, translation and scale invariance. Experiments show that the use of the Convexity Measure of Polygons, combined with classical shape features used for leaf shape characterisation, increases the success rate of leaf shape classification. An average classification rate greater than 95 % was achieved using the multilayer perceptron for all the species in the FLAVIA data set. Good results were also obtained with other classifiers like KNN (90.5 %) and Naive Bayes (88 %). The proposed method outperforms some methods found in the literature. Proposed future works include the design of a new approach for the characterisation of the convexity of object shape inspired by the Convexity Measure of Polygons, as well as the analysis of the shape of colour images using the new design shape characteriser.

References

1. Kulkarni, A.H., Rai, H.M., Jahagirdar, K.A., Upparamani, P.S.: A leaf recognition technique for plant classification using RBPNN and Zernike moments. *Int. J. Adv. Res. Comput. Commun. Eng.* **2**(1), 984–988 (2013)
2. Ayoub, A.B.: The eccentricity of a conic section. *Coll. Math. J.* **34**(2), 116–121 (2003)
3. Hussein, A.N., Mashohor, S., Iqbal, M.: A texture based approach for content based image retrieval system for plant leaves images. In: *IEEE 7th International Colloquium on Signal Processing and its Applications*, pp. 11–14 (2011)
4. Aickelin, U., Bentley, P.J., Cayzer, S., Kim, J., McLeod, J.: Danger theory: the link between AIS and IDS? In: Timmis, J., Bentley, P.J., Hart, E. (eds.) *ICARIS 2003*. LNCS, vol. 2787, pp. 147–155. Springer, Heidelberg (2003)
5. Anant, B., Manpreet, K., Anupam, K.: Recognition of plants by leaf image using moment invariant and texture analysis. *Int. J. Innov. Appl. Stud.* **3**(1), 375–382 (2013)
6. Babu, M., Prasad, S., Rao, B.S.: Leaves Recognition Using Back Propagation Neural Network-Advice For Pest and Disease Control On Crops. *IndiaKisan.Net: Expert Advisory System* (2007)
7. Mezo, B.D., Rabatel, G., Forio, C.: Weed leaf recognition in complex natural scenes by model-guided pairing. In: *4th European Conference on Precision Agriculture* (2007)
8. Bendiab, E., Kheirredine, M.: Recognition of plant leaves using the dendritic cell algorithm. *Int. J. Digital Inf. Wireless Commun. (IJDIWC)* **1**(1), 284–292 (2011)
9. Ji-Xiang, D., Xiao-Feng, W., Guo-Jun, Z.: Leaf shape based plant species recognition. *Appl. Math. Comput.* **185**(2), 883–893 (2007)
10. Zunic, J., Rosin, P.: A new convexity measure for polygons. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(7), 923–934 (2004)
11. Linnus, C.: *Systema natur, sive regna tria natur systematice proposita per classes, ordines, genera, and species*. Lugduni Batavorum (Haak) (1735)
12. Panagiotis, T., Stelios, P., Dimitris, M.: Plant leaves classification based on morphological features and a fuzzy surface selection technique. In: *Fifth International Conference on Technology*, pp. 365–370 (2005)
13. Stephen, G., Forrest, S., Eric, Y., Yu-Xuan, W., Yi-Fang, C., Qiao-Liang, X.: A leaf recognition algorithm for plant classification using probabilistic neural network. *Comput. Sci. Artif. Intell.*, 11–16 (2007)
14. Beghin, T., Cope, J.S., Remagnino, P., Barman, S.: Shape and texture based plant leaf classification. In: Blanc-Talon, J., Bone, D., Philips, W., Popescu, D., Scheunders, P. (eds.) *ACIVS 2010, Part II*. LNCS, vol. 6475, pp. 345–353. Springer, Heidelberg (2010)