PLANTS SPECIES IDENTIFICATION USING COMPUTER VISION TECHNIQUES

HAMROUNI L., AIADI O., KHALDI B. and KHERFI M L.

Univ. Ouargla, Fac, Information Technology and communication
LAGE Laboratory

Abstract: Plants are quite important component in our ecosystem. Botanists need to identify plants type for different targets, for example distinguishing the ones which can be used for medical purposes. Traditionally, botanists identify plants manually by using cellular and biological characteristics, which is, in fact, a tedious and time consuming process. Therefore, designing an automatic system, which is capable to identify the different types of plants, is highly recommended. In this paper, we propose a fully automatic method for leaves classification based on computer vision techniques. Instead of extracting the cellular characteristics of plants, our proposed method recognize the type of the plant from the visual features i.e., characteristics which is extracted from a leaf image. The used features include the leaf length, width and diameter. The proposed method is fully automatic, as it doesn't require any human intervention. In addition, it allows persons who are not familiar with the biology domain to recognize the plants type. To prove the efficiency of the proposed system, we conduct experiments on Flavia dataset which assembles 1907 leaf images of 32 types of plants. Experimental results show promising results and an accuracy of 94% has been reached.

Keywords: plant recognition, morphological features, texture Glcm.

IDENTIFICATION DES ESPECES DE PLANTE A L'AIDE DE TECHNIQUES DE VISION PAR ORDINATEUR

Résumé: Les plantes sont une composante très importante dans notre écosystème. Les botanistes doivent identifier le type de plantes pour différentes cibles, par exemple en distinguant ceux qui peuvent être utilisés à des fins médicales. Traditionnellement, les botanistes identifient les plantes manuellement en utilisant des caractéristiques cellulaires et biologiques, ce qui est en fait un processus fastidieux et long. Par conséquent, la conception d'un système automatique, capable d'identifier les différents types de plantes, est fortement recommandée. Dans cet article, nous proposons une méthode entièrement automatique pour la classification des feuilles basée sur des techniques de vision par ordinateur. Au lieu d'extraire les caractéristiques cellulaires des plantes, notre méthode proposée reconnaît le type de plante à partir des caractéristiques visuelles, c'est-à-dire des caractéristiques qui sont extraites d'une image de feuille. Les caractéristiques utilisées incluent la longueur, la largeur et le diamètre des feuilles. La méthode proposée est entièrement automatique, car elle ne nécessite aucune intervention humaine. En outre, il permet aux personnes qui ne connaissent pas le domaine de la biologie de reconnaître le type de plantes. Pour prouver l'efficacité du système proposé, nous effectuons des expériences sur l'ensemble de données Flavia qui rassemble 1907 feuilles de 32 types de plantes. Les résultats expérimentaux montrent des résultats prometteurs et une précision de 94% a été atteinte.

Mots-clés: reconnaissance des plantes, caractéristiques morphologiques, texture Glcm.

Auteur correspondant : Aiadi O.: o_aiadi@hotmail.com

Introduction

Plants are the backbone of life on the earth and essential resource for human well being. They are considered as the first living organism born on the earth [**1**]. Plants play a decisive role in providing clean air, food, medicine and oxygen. They significantly contribute in protecting and maintaining our environment. In the nature, there exist about 3 million species of plants, each with specific characteristics [**2**]. Identifying plants type is an extremely important task for botanists and scientists from the related fields as well. This is because certain plants are useful for the medical purposes, whereas certain others may be harmful as they can cause diseases or even leading to the death.

Leaves can be used to determine the plant type, as most plant species have unique leaves i.e., a leaf from one type differs from a second from another type in terms of characteristics. Traditionally, botanists and taxonomists adopt different approaches to identify plant species e.g., using molecular biology and cellular characteristics of leaves. However, adopting such approaches for plant identification suffer from serious drawbacks. First, these approaches require performing several critical steps. Second, such approaches are so tedious and time-consuming. Third, only the specialist persons are capable to identify plants, which are, in fact, a serious limitation because identifying plants from no-specialist persons may help in preserving them from the extinction danger. Hence, it is highly recommended, especially with the remarkable development in the computer vision field, to develop an automatic system for plant species identification.

In the literature, the problem of automatic plants identification has been the subject matter of several studies. Nevertheless, it should be mentioned that most studies have focused on identifying plants using their leaves. The general principle of these studies is to take an image for the plant leaf, then extract the visual features i.e., characteristics of the leaf. After that, computer vision algorithms are applied to determine the plant type. For instance, in [**3**] visual features of leaf have combined with Random Forest (RF) and Linear Discriminant Analysis (LDA) classifiers to identify 30 plant species. Similarly, in [**4**], Artificial Neural Networks (ANN) classifier has been trained to identify 12 plant species. [**5**] have focused on identifying the medicinal plants such as herbs, shrubs and trees based on ANN and Support Vector Machine (SVM) classifiers. In our work, however, we use simple yet efficient techniques for plant identification.

In this paper, we develop and design a fully automatic method for plant identification on the basis of leaves images. For a given plant to be identified, we extract the visual features from the leaf image. The visual features include the shape features such as area, perimeter, and diameter. In addition, because leaves of different species may differ in terms of texture, we use the Gray Level Co occurrence Matrix (GLCM) as a texture feature. Then, we use the K-Nearest Neighbor (KNN) classifier to identify the plant type. The proposed method presents multitude advantages: it is fully automatic and it doesn't require any human assistance. Beside, no cellular characteristics are needed for identification, and only the visual features are used. In addition, it allows the ordinary persons (i.e., no-specialist) to identify plant species, which may help in preserving plants from the extinction danger. Moreover, it is capable to perform the identification process in a fraction of second, as shown in the experiments. Experimental results, carried out on the Flavia dataset, have demonstrated the efficiency of our method.

The remainder of this paper is organized as follows section 2 describes shape and texture descriptors presents in our method ,section3 show the experimentation on well known dataset ,section 4 presents conclusion.

Our Method

In this section, we explain our proposed method for leaf identification which is based on morphological and GLCM features. It takes as an input a color image (i.e., image in RGB space), and gives as an output the class to which this image belongs. The method is consists in four main stages namely, preprocessing, feature extraction, features combination and classification stage. Each of these stages will be detailed in the following four sub-sections. Figure1 shows the general scheme of our proposed method.

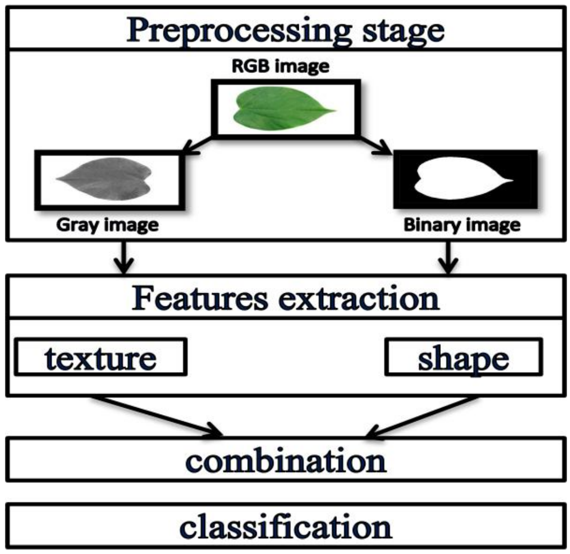

Figure1: General scheme of our method

1. Preprocessing

The preprocessing stage is fundamental in most systems. This stage is, generally, responsible for applying a set of treatment (e.g., noise reduction, rotation, transformation, etc.) on the image before employing it for information extraction. We devote this stage to two main processes, the first one is image minimization while the second one is gray-level image transformation. Thus, this stage gives as output two images, resulted from the original image, which are: gray-level image and binary image.

2. Features extraction

In this stage and having obtained the gray-level and the binary image, we extract a set of shape and texture features, such that, shape features are extracted from the binary image while the texture features are extracted from the gray level image.

2.1 Shape features

Morphological features are shape features that consist in the extraction of the basic geometrical properties [6] of the leaf such as: diameter, area, perimeter, etc.

Fig. 2 shows an example of some geometrical features that can be extracted from a leaf image. In our method, we extract a set of geometrical features including, diameter, area, major axes length, minor axes length, and perimeter. In addition and based on these features, we extract another set of digital morphological features, as described in [6] [7], namely, aspect ratio, Perimeter Ratio of Physiological, Perimeter Ratio of Diameter, rectangularity, narrow factor, circularity and solidity.

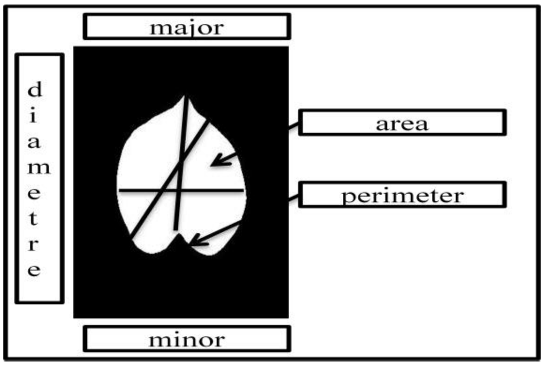

Figure2: Basic geometrical features

- Diameter: (D) is the longest distance between two points of the leaf contour.
- Area: (A) is the number of The pixel count of the leaf area.
- Major axis length: (LP) is the distance between two terminals points orthogonal to minor axis length.
- Minor axis length: (WD) is the longest distance orthogonal to major axis length.
- Perimeter: (P) the number of pixels at the margin of the leaf.
- Aspect Ratio: is defined as the ratio of major axis length LP to minor axis lentgh WP. It is also called as Eccentricity or Slimness. AspectRatio=LP/WP.
- Perimeter Ratio of Physiological length & width:this features is the ratio of perimeter leaf and the sum of major and minor axis lentgh. PRPW=P/(LP+WP).
- Perimeter Ratio of Diameter: it is the ratio of perimeter to the diameter

  PRD=P/D.
- Rectangularity: The similarity between the leaf and a rectangle, given by (Lp*Wp)/A.
- Narrow Factor: the ratio of the diameter D and length Lp, thus D/Lp.
- Circularity: The ratio involving area A of the leaf and the square of its perimeter P given by A/P2.
- Solidity: The ratio between A the area of the leaf and Ach the area of a convex hull given by A/Ach.

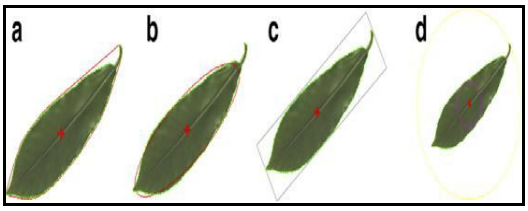

Figure3 [8]: The leaf (a) convex hull, (b) form ellipse, (c) Rectangularity (d) circularity.

2.2 Texture features:

Veins (Fig. 4) are one of the most used characteristics to distinguish between leaves. Therefore, we have tried to employ such an informative characteristic in our method to describe leafs. We, therefore, extract a set of texture features from gray-level leaf images.

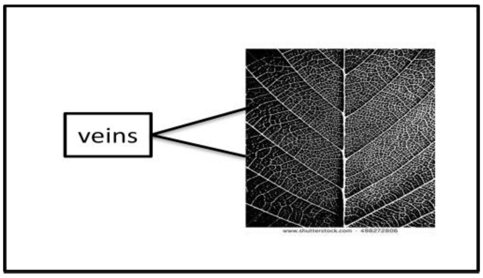

Figure4 : Veins of leaf

These textures features come in form of second order statistical moments extracted from a GLCM [9]. This later is generated co-occurrence frequency of different gray-level pairs within the image. In simpler words, let us suppose that we extract a GLCM *M* from a gray-level leaf image. Then, Each element *M(i, j)* represents the frequency of a gray-level *j* that appears at a given offset $(\Delta x, \Delta y)$ from a gray level *i*. For a given image I and offset (Δx, Δy), the GLCM could be extracted using the next formula (1).

$$M(i,j) = \sum_{p=1}^{N}\sum_{q=1}^{M}\begin{cases}1, & \text{if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0, & \text{otherwise}\end{cases} \quad (1)$$

There are many features that can be extracted from the GLCM including second moment, contrast, correlation, variance, inverse different moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, maximal correlation cofficient and three other mesures of correlation namely f11, f12 and HXY. However, in our method we use only the form most uncorrelated ones which are, Contrast, Homogeneity, Correlation and Energy.

- Contrast: it is stated as the local grey level variation.
- Homogeneity: It measures the uniformity of the non-zero values in the GLCM matrix
- Correlation: It is a measure of linear dependency.
- Energy: It explains how uniform the texture is.

3. Features combinatio

After applying stage 2 (i.e., feature extraction) the results will be two vectors, one for shape and the other for texture features. There are several techniques of features combinations. One may think to establish weights that determine the importance of each feature. However and in order to keep things simple, we simply concatenate these two feature vectors (i.e., shape and texture feature) without assigning any weights. This means that all features are important as the same.

4. Classification

In this final stage, feature vectors, which are extracted then combined in previous stages, are compared in order to identify the appropriate class of a given leaf. This stage is composed of two sub-stages which are learning and recognition.

4.1. Learning stage

The aim of this sub-stage is to train our method so it can distanguish the visiual propoerties of the different leaf class. It takes as input a number of images per each class and then extract feature vectors from each as shown in Fig. 5. Then, the extracte features are served to Knn classifier [10] to generate our trained model. This model will be used later on to identify to which class belongs a given new leaf image.

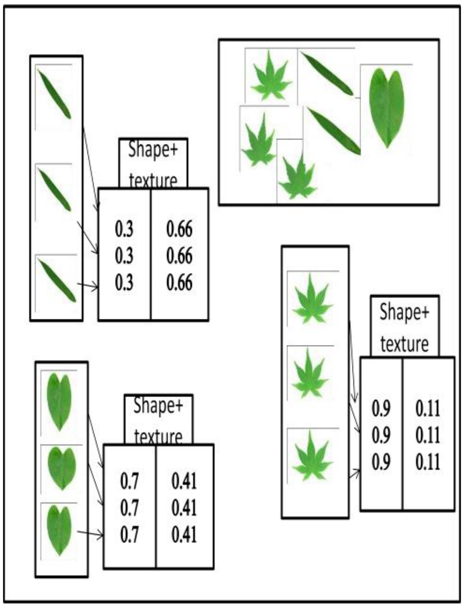

Figure5 : Learning stage

4.2. Recognition

This stage consists in using the pre-trained KNN model in order to find, for a given test leaf sample, the appropriate class. this is done by calculating the similarity between the feature vector of the given test image and those extracted from the training image the similarity could be extracted using the next formula(2).

$$d(\mathbf{x}_r, \mathbf{x}_s) = \left[\sum_{i=1}^{p} c_i (x_{ri} - x_{si})^2 \right]^{1/2}. \qquad (2)$$

Figure below shows recognition stage

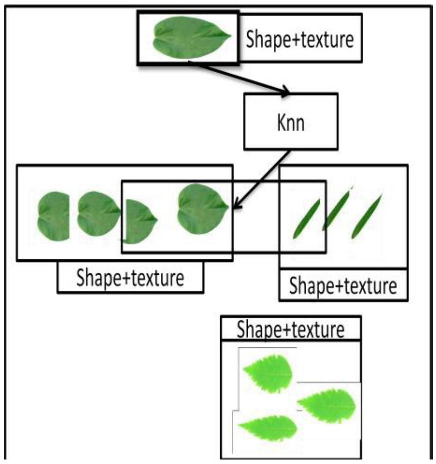

Figure6 : Recognition stage

Experimental results:

In order to evaluate our method we have used Flavia dataset. Flavia is well-known dataset [1] that contains 32 classes with a total of 1907 images. Figure7 shows a representative simple from this dataset.

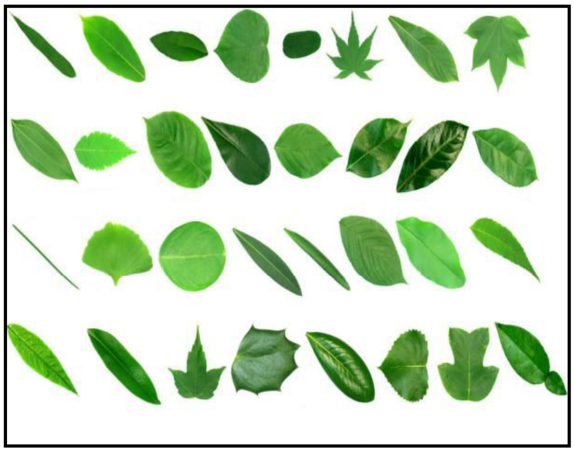

Figure7: Representative samples of each class from flavia dataset

We devote 2/3 of the dataset to train our method, whereas the remaining is devoted to test our method and report its accuracy. Our method has shown high recognition accuracy by recognizing overall 94% of the test

images. It even yields for some class (i.e., species) recognition rate of 100% .
From Table 1, we notice that the combination of two features shape and texture has yielded better recognition rate than one feature. Texture features have outperformed shape features because of the similarity shape of some species.

Feature	Accuracy
Morphological	66%
Glcm	96%
Morphological+glcm	94%

Table: accuracy over all species yielded by each combination

Conclusion

Plants present an essential resource for human life such as food and medicine. Therefore an automatic system is needed to identify different plant species. In this paper we have proposed a method based on morphological features in order to capture shape properties of plants. Because of some species are shapes alike, we have included also GLCM texture feature to capture internal structure of leaf veins. Experimental evaluation shows that our method yields excellent results with overall accuracy of 94% and 100% for some species and in quickly time.

References

[1] **Arunpriya C., et al. An Efficient Leaf Recognition Algorithm for Plant Classification Using Support Vector Machine (2012).** Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering, March 21-23.

[2] **Harish, B.S. ; Hedge, A. ; Venkatesh, O. ; Spoorthy, D.G. ; Sushma, (2013).**Classification of plant leaves using Morphological features and Zernike moments" Advances in Computing, Communications and Informatics (ICACCI), International Conference on DOI: 10.1109/ ICACCI.2013.6637459, Page(s): 1827 – 1831.

[3] **Esraa Elhariri 1,*, Nashwa El-Benday2,*, Aboul Ella Hassanien3,* (2014)** .Plant classification system based on leaf features. 9th international conference on computer engineering and systems (ICCES), pp 271–276.

[4] **G. Kaur and G. Kaur (2012)** Classification of Biological Species Based on Leaf Architecture, International Journal of Engineering Research and Development, Vol 1, Issue 6, pp. 35-42

[5] **B. S. Anami, S. S. Nandyal and A. Govardhan, (2010).** A Combined Color, Texture and Edge Features Based Approach for Identification and Classification of Indian Medicinal Plants, International Journal of Computer Applications, Vol. 6, No. 12, pp. 45-51.

[6] **S. G. Wu, F. S. Bao, E. Y. Xu, Y. X. Wang, Y. F. Chang and Q. L. Xiang, (2007).** A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network. IEEE International Symposium on Signal Processing and Information Technology, Giza, pp. 11-16.

[7] **Pedro F. B. Silva, Andre R.S. Marcal and Rubim M. Almeida da S ilva(2013)**. Evaluation of Features for Leaf Discrimination , In Image Analysis and Recognition, Springer Lecture Notes in Computer Science, vol. 7950, pp. 197-204., Springer Berlin Heidelberg.

[8] **Du, J.-X., Wang, X.-F., & Zhang, G.-J. (2007).** Leaf shape based plant species recognition. Applied Mathematics and Computation, 185883–893.

[9] **R. M. Haralick, K. Shanmugam, and I. Dinstein(1973).**Textural features for image classification, IEEE Trans. System Man. Cybernetics, vol. SMC- 3, pp. 610 621.

[10] **Buttrey, S.E., Karo, C. (2002)** Using k-nearest-neighbor classification in the leaves of a tree. Computational Statistics and Data Analysis 40, 27–37.