Efficient Tool for the Recognition of the Leaves of Plants

Abdelhadi Adel [1] and Mouss Leila-Hayet [2]

[1] **Department of Industrial Engineering, University of Batna, Algeria**

[2] **Department of Industrial Engineering, University of Batna, Algeria**

Abstract

This work appears in pattern recognition in the agronomic domain, especially for the identification of the leaves of plants, while using the adaptive technique of neuronal networks.
In this article, we will expose our tool; which is intended for two categories of specialists, the first consisting of researchers in the field of botany, as the second, so all scientists, who may use this work in their own applications.
We will expose also, the capacities of generalization of the neuronal networks and their implementation to our problem.

Keywords: *pattern recognition, neuronal networks, classification, leaves of plants, coding, meadow treatment, training, analysis.*

1. Introduction

The problem of classification of the leaves of plants is related in large part to the problem of pattern recognition (PR), which represents the first step in a long process of understanding of our universe in the framework of man machine communication. The classification of images is a difficult task for several reasons; first, forms belong to a physical world with digital transcription is very complex, because the absence of sensors suitable for all situations, then the nature forms and their appearance varies from one sample to another, multiplying the dimensions of the representation space and lengthens the time of decision.
The act of classifying plants is an essential task for any person interested in botany, because no one can claim to know all about the different species of plants, because of their high number and great similarities in their shapes, colors, etc....
Each time you encounter a new type of plants, we must devote an enormous effort of research in encyclopedias and dictionaries, to determine its species.
Most work in this sense are not characteristic of automation (determine the type of plant with access to the properties of its leaves), This is due to the absence of conventional criteria, helping to classify plants.
E. Hans. And M. Guédès [5] have developed a manual that contains a list of plants (about 170 species) classified according to the possibility of their edible fruits, and the general shape of their leaves.
A. Michell [8] incorporates in his book 250 types of trees classified according to characteristics of the shape of their leaves.
Our work aims to avoid this lengthy procedure, and can be considered as an efficient tool helping to identify plant species, using samples of their leaves.
The objectives of this work are:
- Find an efficient tool, quick and intelligent helps to distinguish between types of plants deferens.
- Build a database containing information on plants and their characteristics, with the possibility of updating at any time (for new species).

Finding a solution to our problem, with the ability to generalize to other problems in other applications, so it may be a start for other work, it suffices to choose only the features of the new field studied.

2. Application of Neural Networks for Pattern Recognition of the Leaves of Plants

First, as this is a problem of pattern recognition (PR), it appears obvious to present the classical approach to this problem [1] which is illustrated in Figure 1.

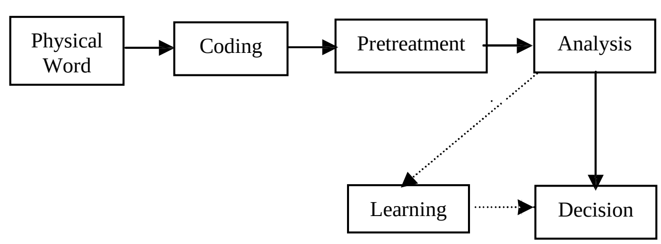

Fig. 1 General Schematic of System of Pattern Recognition.

We will start with a simple explanation of each step in developing the process of recognition:

2.1 Physical world

Is an infinite dimensional analog called space form. The objects in this space are described in different ways with a multitude of properties which would be difficult to consider each of the PR.

2.2 Coding

Is a digital conversion of the physical world to a continuous discrete digital world. The latter also called representation space.

2.3 Pretreatment

The pre-treatment step is to select in the space of representation information necessary to implement and which characterizes the shape. This selection is often through the elimination of noise, for data standardization and the elimination of redundancy.

2.4 Analysis

At the analysis step, the techniques of RF calculate a number of characteristics or parameters. These parameters correspond to measures of geometric nature, characteristics, and statistics.

2.5 Learning

Is a key step in the chain of recognition. Its role is to clarify the decision with a priori knowledge on the forms. Starting from specific forms, learning attempts to define reference models or to characterize the classes of decision.

2.6 Decision

Decision or classification is the recognition step itself, its role is to identify the shapes tested from learning achieved. The method of decision is' displayed 'in learning, which means that the criteria used for learning are the same for the comparison.

The two phases, learning and decision phases which are networks of neurons involved. The network we have chosen is a multilayer network, has the following characteristics [1]:

- The inputs and outputs are the leaves of plants and species deferens.
- The learning algorithm is reversing the propagation of error.
- The number of entries depends on the number of parameters calculated in the analysis phase.
- The number of outputs depends on the number of classes to model plants.

3. Characteristics of Forms

The forms have certain features (perimeter, area, etc...), which provide sufficient information for the applications of pattern recognition.
These features can be used as descriptors of objects, because they may represent the forms in a small vehicle. The characteristics of shapes can be generated from the binary image or original. The characteristics of shapes can be grouped into two groups: Those border regions and those of [2].

4. Extraction of parameters of leaves of plants

We regard the leaves of plants to extract the parameters characterizing the shape and color variation. Broadly speaking we have proposed two major classes of parameters:

- Parameters measured on the form
- Parameters measured on the color

It should be noted that the value of one parameter separated does not make a distinction between the vas deferens leaves (i.e.: you can find two leaves deferens with values equal of the same coefficient).

4.1 Compactness(C)

$$c = \frac{Su}{Pe} \qquad (1)$$

Where:
Su: represents the surface (equal to the number of pixels located within the border of the leave of the plant).
Pe: represents the perimeter (equal to the number of pixels along the border of the leave of the plant).
This parameter gives the relationship between the surface and the perimeter, because there is a perimeter surrounding a large area (Figure 2).

Fig .2 Efficiency of compactness1.

And the same perimeter can enclose a small area (Figure 3).

Fig .3 Efficiency of compactness2.

We fixed Pe (perimeter)

If Su ↗ ⇒ C ↗

If Su ↘ ⇒ C ↘

we fixed Su (surface)

If Pe ↗ ⇒ C ↘

If Pe ↘ ⇒ C ↗

4.2 Elongation (Along)

$$Elong = \frac{Leng}{Width} \qquad (2)$$

Where:

Leng: represents the length of a leaf the distance between the two ends Ext1, Ext2.

Width: represents the width of a leaf (between the two ends Ext3, Ext4 (Figure 4).

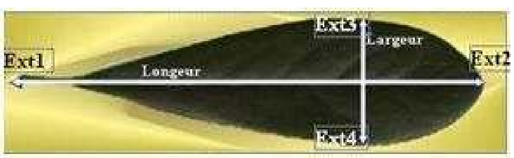

Fig .4 Elongation.

This parameter gives the relationship between the length and width of a leaf, so it distinguishes between the broad-leaved and narrow leaves.

Figure 5 shows that the value of this parameter is fixed to the leaves similar in terms of form.

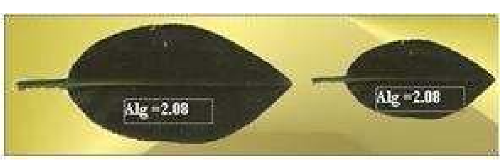

Fig. 5 Attaching Elongation for similar forms.

4.3 Level parabolic (LP)

$$LP = \frac{Su}{Srect} \qquad (3)$$

Where:

Su: represents the surface of a leaf.

Srect: represents the surface of rectangle Drawn from the 4 ends of the leaf (Figure 6).

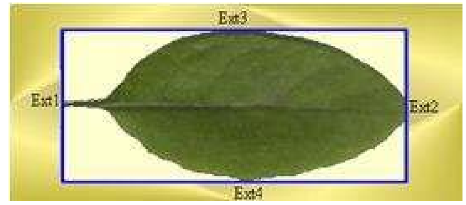

Fig. 6 Level parabolic.

This parameter expresses the degree of parabolic of leave, so we can see quantitatively the distances between the ends symmetrical.

4.4 Relativity (R)

$$R = \frac{D1}{D2} \qquad (4)$$

Where:

D1: distance between the projection of ext3 on the right (D) (linking Ext1 and Ext2) and Ext1 (Figure 7).

D2: length of the leaf.

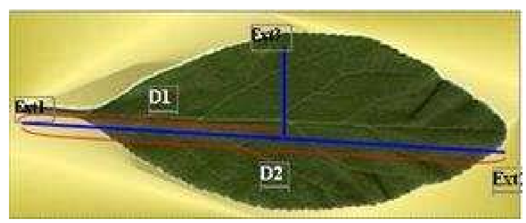

Fig. 7 Relativity.

This parameter allows distinguishing between the leaf area which is parabolic near the end of their supports, and those whose area is below (Figure 8).

Fig. 8 Efficiency of LP.

4.5 Concentration of intervals of distance (CID)

This parameter is divided to ten parameters, each of represents the number of contour point whose distance between them and the right (ext1, ext2) is located in an interval (from ten intervals defined above (Figure 9).

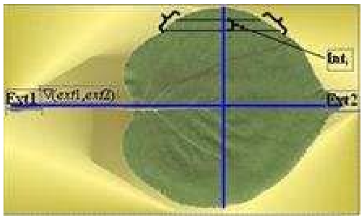

Fig. 9 Concentration of intervals of distance.

How is taking the choice of the ten intervals?

Assuming that F1 has a leaf with a width g_1, the length of interval ***int$_i$*** $L = \frac{g1}{20}$ all intervals to be chosen are

$$Int_1 = [0,L]$$
$$Int_2 = [L,2*L]$$
$$\ldots\ldots\ldots$$
$$Int_{10} = [9*L,g1].$$

This parameter gives us an idea of length of support of the leaf relative to its total length.
There is a strong relationship between this parameter and the level Parabolic (LP) (an increase of LP affects the following way: the value of the parameters under which corresponds to the first interval increases, and conversely for what they correspond to the last interval).

4.6 Gray Level Medium (GLM)

To better understand this parameter how Windows manages color systems on the types Bitmap, Windows API provides the structure TRGBtriple (3 bytes), which describes the three colors: Rgbtred, Rgbtgreen, Rgbtblue. Each pixel of the image has a gray level represented by these three colors.
This parameter is composed of three parameters R-CM, CM-G, CM-B) Each of them represents the medium of a color at all points of image.
The color plays a very important role in the classification of leaves of plants. We use the term to express the average change of colors in each leaf studied (Figure 10).

Fig. 10 Efficiency of GLM.

4.7 Range of Color (RC)

The information carried by this parameter is the bound of the interval of a leaf color varies (Figure 11).

Fig. 11 Efficiency of RC.

5. Tool RLP (Recognition of the Leaves of the Plants)

The tool RLP aims to reduce these difficulties, particularly for those who are changing the research environment frequently, after each move they meet new plants ever seen, so instead of browsing indexes and the pages of dictionaries, just take some samples of leaves, and bring them to the RLP, and to determine the type corresponding to the plant.

5.1 General architecture of RLP

General architecture of RLP is illustrated in Figure 12:

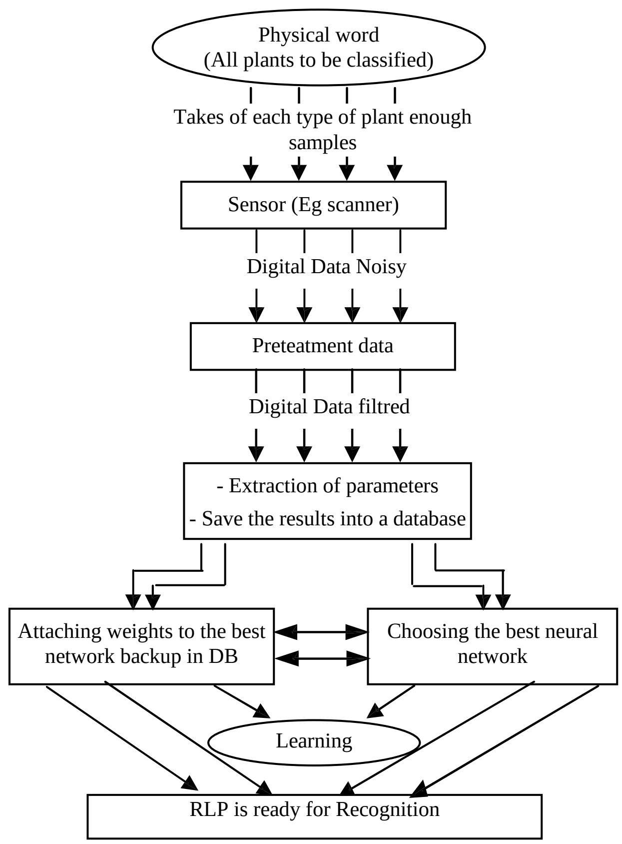

Fig .12 General Architecture of RLP.

ISSN (Online): 1694-0814
www.IJCSI.org

6. Experimentation and Results

6.1 Characteristics of samples

Note that RlP is able to recognize the leaves of plants whose learning has been done. In addition, samples must satisfy certain constraints:

- All samples must be representative for the corresponding types, i.e. take samples at each age of the leaf.
- Neglecting the pathological cases (cases where the sheet does not have the same characteristics of most of the leaves).
- Support of the sheet must be taken into account because it represents an essential characteristic.

6.2 Choosing the best neural network

The process of choosing the best network is the following:

- Determine the number of iterations.
- Validate use.
- Run learning for all possible combinations (about: learning coefficient, number of hidden layers, number of neurons in the hidden layers) at the end of learning in a network in its stock with the error.

The best network is the network that corresponds to the smallest error.

6.3 Application Example

We test RLP on an image is 100 species of plants, each of which is divided into two classes:

- Class leaves learning to drive the selected network (from 4 to 8 leaves per plant).
- Class leaves test: to test the results of learning (4 leaves for each plant).

The neural network shows good performance of associations on all test sheets, especially sheets homogeneous in terms of form.

It should be noted that the use of semi distributed coding affects the speed of convergence of the network (approximately 3 times faster than the conventional coding), the two following graphs shows this influence.

The two following figures show the variation of error for the semi distributed coding and coding standard.

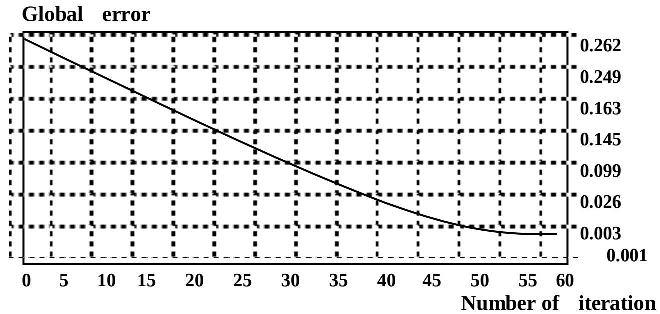

Fig. 13 Graph of error (semi distributed coding)

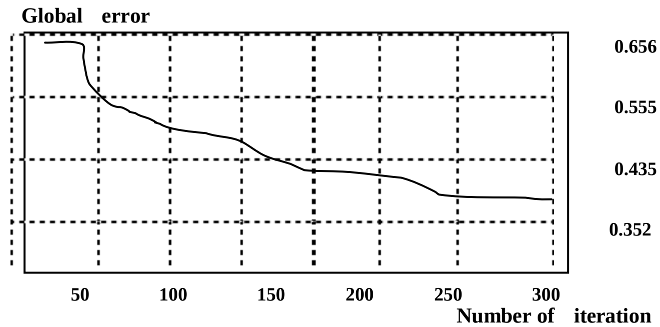

Fig. 14 Graph of error (conventional coding)

6.4 Results

All leaves training is presented to the network (in the learning phase) 100,000 times (number of iterations).

During these iterations the network has a strong tendency to reduce the total error, especially the local error in the case of semi distributed coding (0.7 after a few iterations and 10-8 at the end of learning).

The network is able to respond correctly to 83 plants species among the 100 tested in the case of semi-distributed coding, giving a success rate close to 85%.

The number of plants species recognized by the network in the case of conventional coding, no more than 70% leaves (70% success rate).

7. Conclusion

Since we began this work, our goal was to achieve a complete product, intended for two categories of specialists, the first consisting of researchers in the field of botany, as the second, so all scientists, who may use this work in their own applications.

We were able to find parameters, most of which are mathematics, which vary depending on the measures calculated on the form, and which allows to classify the plants, each in a category; the calculation of these parameters is done in a way Automatic image of the paper presented to the program without any user intervention.

ISSN (Online): 1694-0814
www.IJCSI.org

The method of neural networks was the tool of recognition and classification, and as the solution of choice of network to implement.

We were able to exceed 85% of success rate for all species of learning, in the case of semi-distributed coding, 70% for conventional coding. This rate, which may be considered satisfactory, can be traced back to a near certainty if we add other parameters, such as the chemical and cellular leaves.

As prospects for this work are:

- Add new parameters, such as the chemical and cellular constituents of the leaves of plants with their PH, to achieve a ranking in the same species.
- Implement our best neural network on a parallel machine to exploit the concept of parallelism offered by neural networks.

References

[1] A. Belaid, Y. Belaid, "Reconnaissance de Formes: Méthodes et application", Paris, France, 2002.
[2] M. Boumehrez, "Utilisation des réseaux de neurones pour la reconnaissance des caractères", mémoire de fin d'étude, Université de Batna, Algérie, 2005.
[3] G. Deryfus, "Les réseaux de neurones formels, revue", Flux supelec, pp 13-17, 2008.
[4] A. Ghodbane, "Contrôle des systèmes non linéaires par réseaux de neurones", mémoire de fin d'étude, Batna, Algérie, 2005.
[5] E. Hans, M. Guèdes, "Le monde des plantes ", édition Oberville, 2000.
[6] R. Hush, B. Home, "Progress in supervised neural network", revue: IEEE signal processing magazine, 2009.
[7] K. Kara, "Application des réseaux de neurones pour l'identification des systèmes non linéaires", 2007.
[8] A. Michell, "Caractéristiques des arbres", édition Eyrolles, 2005.
[9] D. Rumelhart, "Learning representation by back-propagating errors", 2006.
[10] P. Simon, "Artificial neural systems", university of sandiego, USA, 2000.
[11] M. Wickelgren, "Context sensitive coding associative memory", 2009.

Abdelhadi Adel received the magister degree in 2004 from the Department of Computer Science, University of Batna, Algeria. He is currently an assistant professor at the University Center of Khenchela, Algeria. He is currently a Doctoral student in the Department of Industrial Engineering, University of Batna, Algeria. His research interests include preventive maintenance, multi-agent system, classification, diagnosis and artificial immune system.

Mouss Leila-Hayet is professor in the Department of Industrial Engineering, University of Batna, Algeria. Here research interests include maintenance, diagnosis, Scheduling and Production.