Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp326-331)

Recognition system for wooded species of Canary Laurisilva from its contour using kernel of Fisher

CARLOS M. TRAVIESO[1], JUAN C. BRICEÑO[2], JESÚS B. ALONSO[1], MIGUEL A. FERRER[1]
[1]Departamento de Señales y Comunicaciones
Universidad de Las Palmas de Gran Canaria
Campus de Tafira, Edificio de Electrónica y Telecomunicación, E-35017 Las Palmas de Gran Canaria
SPAIN
[2]Escuela de Ciencias de la Computación e Informática
Universidad de Costa Rica
Sede "Rodrigo Facio Brenes", Montes de Oca, Código Postal 2060, San José.
COSTA RICA
http://www.gpds.ulpgc.es

Abstract: - In this work it has been developed a novel system of automatic recognition for trees leaves, based on a hybrid classification method by means of a Support Vector Machine (SVM), using the Fisher kernel, and calculated from a Hidden Markov Model (HMM). An angle sequence has been used as an element of parameterization, extracted from the leaves contour, making such sequence invariable by rotation, movements and size. The application of these algorithms has been focused on the implementation of a recogniser of leaves of endemic trees from the Canary Islands, in particular, 16 different species of Canary Laurisilva. Obtained successful rates have been higher than 99.9%, according to the number of employed leaves in the training process.

Key-Words: - Pattern Recognition, Artificial Intelligent, contour detection, parameterization, classification, HMM, kernel of Fisher and SVM.

1 Introduction

The use of the information technologies and communications is becoming more present in diverse fields of our society. The field of botany is not an exception. This work intends to be itself, a tool that helps biologists with their work.

Treating this theme in detail, in the vegetable kingdom there are catalogued more than 350.000 species [1][2]. The variety of living beings is so high, that a detailed classification for each one has become essential. This task is performed by pattern recognition and, is in this environment, where this work fits well.

Biologists are in charge of cataloguing the enormous amount of vegetable species in the planet. This splendid biodiversity makes it difficult to know accurately which species we have, even for scientists. This is why it is required to consult with specialized experts in the vegetable families, to know which specie the studied leaf belongs to. Nevertheless, there not always an expert nearby in this field to tell us what plant we are studying.

Therefore, this work is born from the need to count with an efficient and fast method to get to classify diverse vegetable species, based on signal processing techniques [3]. To give answer to this lack, it would be desirable to develop a digital processing program that, by means of a series of specific designed tools for image processing, managed to identify different vegetable species.

By this way, the laborious consultation in encyclopaedia and specialized literature would be avoided in order to find out what plant is the subject of study. In just a few seconds it would be possible to know what plant is being observed.

Hence, the work methodology to follow has been based on the conjunction of different ideas. On one hand, working with leaves contour using the information provided. On the other hand, using the Hidden Models of Markov (HMM) [2][4] to calculate the Fisher kernel [5]. These data will then be classified by means of the support vector machines (SVM) [6][7], to eventually obtain a model through supervised learning that will give a certain success value for the recognition of the tree leaves. A plan of the system proposed can be seen in figure 1.

The structure of this paper is established after this introduction in the section that includes the creation of the database. In the third section the leaves contour is calculated. In the next section, the Hidden Markov

Model (HMM) is introduced as a transformed element to obtain the Fisher kernel. The fifth section is dedicated to the classifier study, experiments and results, and finally the seventh section shows the conclusions.

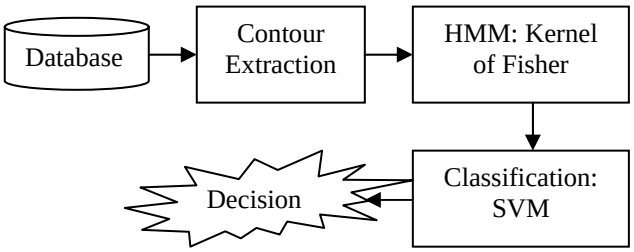

Fig 1. Blocks diagram of the proposed system.

2 Created database

In order to create a recognition system of different vegetable species it is necessary to build a database. This database should contain the samples of the different species of study. The number of samples will be large enough to, first train the classifier with guarantees and second, test this classifier to assess the results obtained. On top of this, the amount of chosen samples, for each vegetable species must cover the largest amount of shapes and structures that this unique specie can take. In this way, a robust study of the different vegetable species is ensured.

Attending to this reasoning, the sample collection was made at different times of the year, trying in this way to cover all the colours and shapes that the

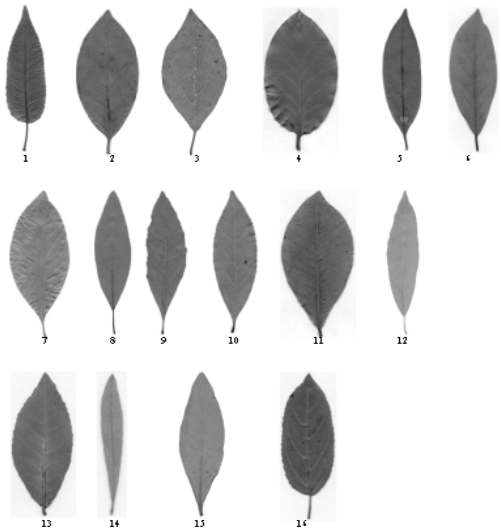

Fig. 2. Images of the 16 varieties of canariensis laurisilva considered for the present study. Images are presented regardless of size.

leaves take throughout the four seasons. Besides, a special attention was made to reject those samples that were degraded so that the selected samples were in good conditions.

The species that were chosen have the singularity of being endemic to the Canary Islands, which can be found in low areas (Mocán) and the humid mountainous areas of Laurisilva and Fayal-Brezal [8].

The database is composed of 16 different species, with 75 samples each one. The images that form the database has been stored in a grey scale using a "jpeg" format (Joint Photographic Experts Group) with Huffman compression. The images have been digitalised to 300 dpi, with 8 bit accuracy [8].

3 Contour Detection

We have considered just the leaf perimeter. This image is considered without its petiole that has been extracted automatically from the shadow image.

Leaves are scanned fixed on white paper sheets, placed more or less on the center, upward (petiole down) and reverse side to scan.

Border determination as (x,y) positioning perimeter pixels of black intensity, has been achieved by processes of shadowing (black shape over white background), filtering of isolated points, and perimeter point to point continuous follow.

3.1 Perimeter interpolation.

As shown in table 1, perimeter size variability induces us to consider a convenient perimeter point interpolation, in order to standardize perimeter vector description.

For an interpolating process, in order to achieve reconstruction of the original shape, we may use any of the well known algorithms as mentioned in [9][10][11], but a simple control point's choice criterion in 1-D analysis allows for an appropriate performance ratio on uniform control point's number and approximation error for all individuals of all varieties studied.

The general idea, for such choice, is to consider (x,y) positional perimeter points as (x,F(x)) graph points of a 1-D relation F.

Consideration of y coordinate as y = F(x) is done, because of the way leaves images are presented in our study: leaves have been scanned with maximum size placed over x ordinate.

For a relation G to be considered as a one-dimensional function, there is need to preserver a correct sequencing definition (monotonic behaviour).

That is: A graph,

$$G = \{ i = 1..n, (x_i, y_i) / y_i = f(x_i) \} \quad (1)$$

is the description of a function f if ordinate points $x_i, i=1..n$ must be such that: $x_i < x_{i+1}, i=1..n-1$.

We consider then the border relation F as a union of piece like curves (graphs) preserving the monotonic behaviour criterion, i.e.

$$F = \bigcup_{j \in J} G_j \quad , \qquad (2)$$

where: $G_j \subseteq F, \forall\, j \in J$

and $G_j = \{\alpha_j \in J_j, (x_{\alpha_j}, y_{\alpha_j}) / y_{\alpha_j} = f_j\}$,

For convenient sets of index J, J_j and restriction functions $f_j = f_{|\{x_{\alpha_j} | \alpha_j \in J_j\}}$, such that the next point following the last of G_j is the first one of G_{j+1}. G_j graphs are correct f_j functions descriptions.

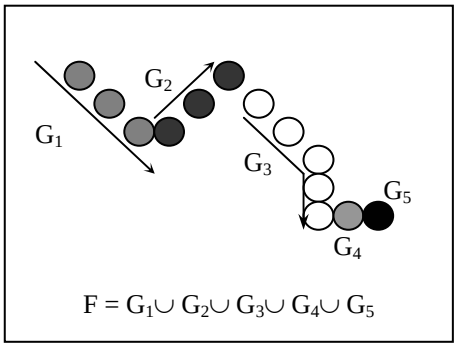

Fig. 3. Example of an F relation decomposed in graphs with a correct function description.

Building the G_j sets is a very straightforward operation:

- Beginning with a first point we include the next one of F.
- As soon as this point doesn't preserve monotonic behaviour we begin with a new G_{j+1}.
- Processes stop when all F points are assigned.

In order to avoid building Gj reduced to singletons, as show in figure 3 (G_4 and G_5) the original F relation may be simplified to preserve only the first point of constant x ordinate series.

Afterwards, spreading of a constant number of points is done proportional to the length of the G_j and always setting in it is first one.

The point's choice criterion mentioned before allows, in two-dimensional interpolation, for taking account on points where reverse direction changes take place. Irregularity, of the surface curve, is taken into account with a sufficient number of interpolating points, as done in the uniform spreading way. Results on table 1 allows for comparison between choice of control points with the criterion motioned before and the uniform one. Such results show the benefit of choosing control points with the monotonic criterion instead of the uniform one.

Class	Mean size	Mean error	
		Uniform	Monotonic
01	2665.6	9.1474	2.0226
02	1885.1067	3.5651	0.43655
03	2657.68	11.0432	5.3732
04	2845.8133	31.6506	2.8447
05	1994.68	1.8569	0.42231
06	2483.04	0.4425	0.71093
07	2365.2667	9.711	0.68609
08	3265.48	0.4753	0.49015
09	2033.2267	19.7583	3.4516
10	2258.2533	3.9345	2.4034
11	1158.9867	5.4739	1.0286
12	1934	1.3393	0.40771
13	1183.4	1.2064	0.39012
14	981.4	0.2752	0.23671
15	3159.08	11.575	8.8491
16	1973.3733	47.4766	6.6833

Table 1. A comparative table of mean error, obtained from a uniform criterion of control point selection and the monotonic way

In table 1, sizes are given in number of pixels. The 1-D interpolation has been perform using 359 control points, with spline, lineal or closest interpolated point neighbourhood, depending on the number of control points present in the decomposed curve (ie. each G_i). As a reference at 300 dpi a crayon free hand trace is about 5 to 6 points wide.

Table 1 also shows size variability of the different varieties ranging in mean, between 981 pixels for class 14 to 3255 for class 8. With 359 points chosen with the monotonic criterion, all perimeter point vectors have a standard size and errors representation is negligible.

Due to perimeter size variability inside a class, for example in class 15 ranging between 2115 points to 4276 with a standard deviation of about 521, coding of (x,y) control perimeter points have been transformed taking account for size independence.

Considering the following definitions:

Γ the set of n, a fixed number, of control points, $\Gamma = \{X_{i=1..n} / X_i = (x_i, y_i)\}$ Where (x_i,y_i) are point coordinates of control perimeter points.

C_0 the central point of the Γ set: $C_0 = (1/n)(\sum_{i=1..n} x_i, \sum_{i=1..n} y_i)$, $/ (x_i, y_i)_{i=1..n} \in \Gamma$,

$\beta_i = angle(C_0 X_i X_{i+1}), \alpha_i = angle(X_i C_0 X_{i+1})$

angles defined for each interpolating points of Γ. An example is shown in figure 4.

Sequences of (x_i,y_i) positional points are then transformed in sequence of (φ_i, β_i) angular points.

The choice of a starting and a central point accounts for scale and leaf orientation. Placement of

both points sets the scale: its distance separation. Relative point positioning sets the orientation of the interpolating shape. Given a sequence of such angles: α_i and β_i, it's then possible to reconstruct the interpolating shape of a leaf. Geometrical properties of triangle similarities make such sequence size and orientation free.

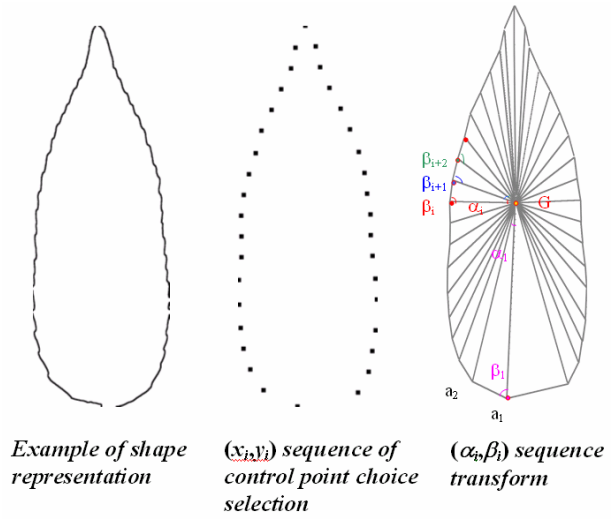

Fig. 4. Example of an angular coding for a 30 control points selection.

4 HMM: the Fisher kernel.

A HMM is the representation of a system that, for each value that takes a variable t, called time, it is found in one and only one of N possible states and declares a certain value at the output. It can tell that a HMM carries associates two stochastic processes: one hidden (not observable directly) associate with the probability of transition between states; and another observable one, associate with the probability to obtain each one of the possible values at the output, and that depends on the state in which the system be found [2][4].

A HMM is defined by: N number of states, M the number of different observations, A(N,N) is the probabilities matrix of transition from a state to another, π(N,1) is the vector of probabilities that the system begin in a state or another; and B(N,M) is the probabilities matrix of that being in each one of the possible states, it be produced each one of the possible observations [2][4].

They exists certain HMM called "from left to right" or Bakis, that are especially appropriate for the sequences. These Bakis HMM turn out to be especially adequate to shape the contour of the leaves because the transition through the states is produced in a unique direction, so that in the extent of time elapsing, the model advances for the states that we well expects. This provides, to this type of models, the ability to keep in mind the order in which the observations they are produced, as well as the temporary distance among the most representative changes.

Finally, it is proposed the transformation that provides the HMM probabilities, relating to the approach of the Fisher score [6]. With this goal, it intends to unite the probability given by the HMM to the given discrimination of kernel of the SVM, whose tie of union is this Fisher score. This score calculates the gradient with respect to the parameters of HMM, in particular, on the probabilities of emission of a vector of data x, while it is found in a certain state $q \in \{1,..,N\}$, given by the matrix of symbol probability in state q ($b_q(x)$), just as it is indicated in equation 3;

$$P(x/q,\lambda) = b_q(x) \tag{3}$$

If it is realized the derivative of the logarithm of the above probability, with the purpose to calculate its gradient, it is obtained the kernel of Fisher, whose expression is given by [6];

$$\frac{\partial}{\partial P(x,q)} \log P(x/q,\lambda) = \frac{\xi(x,q)}{b_q(x)} - \xi(q) \tag{4}$$

where in [6], it has been found approximations, and calculations for the above equation.

Besides, $\xi(x,q)$ represents the number of times, that it is localized in a state q, during the generation of a sequence, emitting a certain symbol x [4][6]. And $\xi(q)$ represents the number of times that has been in q during the process of generation of the sequence [4][6]. These values are obtained directly and of form effective, from the forward backward algorithm, applied to the HMM [4].

The application of this score (U_X) to the SVM, comes given by the expression of the equation 4, utilizing the techniques of the natural gradient, from the following equation 4;

$$U_X = \nabla_{P(x,q)} \log P(x/q,\lambda) \tag{5}$$

where U_X define the direction of maximum slope of the logarithm of the probability to have a certain symbol in a state.

5 SVM Classification

The reasons for choosing SVM are because it is a system based on geometric properties and that can use a great number of samples for training process.

SVM works with a system bi-class, that is, only it is capable of discriminating between two different classes, therefore we have built a multiclass system. The basic idea consists of coaching the system to obtain two sets of vectors (in two dimensions correspond with points) that represent classes to differentiate. Subsequently, the separating hyperplane is calculated, H (in two dimensions is a straight line) between these two sets. The pertinent points al hyperplane should satisfy the following equation [7][12]:

$$w \cdot x + b = 0 \quad (6)$$

where:

- w is normal to hyperplane
- b/||w|| is the perpendicular distance from the hyperplane to origin
- ||w|| is the Euclidean norma of $\overline{w}$
- b is the independent term
- x a point conteined in the plane.

Also other two hyperplanes are defined, H_1: $x_i \cdot w + b = 1$ and H_2: $x_i \cdot w + b = -1$, that are the ones that contain the vectors (or points) more nearby. These vectors receive the name of support vectors. The distance between planes H_1 and H_2 is known like the margin.

As already it has been commented, the support vector machines (SVM) are based on a system bi-class, that is, only two classes are considered to the input of the classifier. Therefore, strategies have been developed, oriented to the classification by means of SVM in multiclass systems, it is defined as one front remainder [12].

In the SVM, it is calculated the separate between patterns, by means of the calculation of the natural distance among the score of two sequences X and Y (it generates H_1 y H_2) [12];

$$D^2(X,Y) = \frac{1}{2}(U_X - U_Y)^T F^{-1}(U_X - U_Y) \quad (7)$$

where F is the information matrix of Fisher, and is equivalent to the matrix of covariance of the vectors U_X and U_Y.

Finally, different types of functions can be used for classify the Fisher kernel, particularly we are going to work with a lineal and gaussian kernel, although this last produces better discriminations among pattern, being shown this kernel in the following equation [12];

$$K(X,Y) = e^{-D^2(X,Y)} \quad (8)$$

6 Experiments and Results

The experiments have been based on the parameters calculated of the detection of the contour. From these we have calculated different Fisher kernels, varying the number of states of the HMM, and using between the 5% and the 30% of the samples of our database.

Each experiment has been repeated in five times, to eliminate peak rates. And therefore, the results are shown in average and standard deviation. In the following tables, the results obtained between 20 and 80 states are shown.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	7×10^{-7}	99.51% ± 0.60	99.56% ± 0.66
10 (8 hojas)	6×10^{-7}	100% ± 0	100% ± 0
15 (12 hojas)	7×10^{-7}	100% ± 0	100% ± 0
20 (15 hojas)	3×10^{-7}	100% ± 0	100% ± 0
25 (19 hojas)	1×10^{-7}	99.87% ± 0.30	99.84% ± 0.29
30 (23 hojas)	3×10^{-6}	99.95 % ± 0.11	100% ± 0

Table 2. Results from SVM using a HMM with 20 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	2×10^{-7}	99.35% ± 1.08	99.38% ± 1.00
10 (8 hojas)	2×10^{-7}	99.89% ± 0.12	99.92 %± 0.10
15 (12 hojas)	7×10^{-7}	99.98% ± 0.09	99.96% ± 0.04
20 (15 hojas)	7×10^{-7}	100% ± 0	100% ± 0
25 (19 hojas)	9×10^{-8}	100% ± 0	100% ± 0
30 (23 hojas)	7×10^{-7}	99.98% ± 0.05	99.98% ± 0.05

Table 3. Results from SVM using a HMM with 30 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	8×10^{-7}	99.81 % ± 0.20	99.86% ± 0.15
10 (8 hojas)	7×10^{-7}	99.94% ± 0.08	99.96% ± 0.08
15 (12 hojas)	7×10^{-7}	99.94% ± 0.09	99.92% ± 0.08
20 (15 hojas)	7×10^{-7}	99.96% ± 0.09	99.96% ± 0.09
25 (19 hojas)	8×10^{-7}	100% ± 0	100% ± 0
30 (23 hojas)	7×10^{-7}	100% ± 0	100% ± 0

Table 4. Results from SVM using a HMM with 40 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	$5.\times10^{-7}$	99.93% ± 0.11	99.93% ± 0.11
10 (8 hojas)	7×10^{-7}	99.90% ± 0.12	99.90% ± 0.12
15 (12 hojas)	7×10^{-7}	100% ± 0	100% ± 0
20 (15 hojas)	5×10^{-7}	99.96% ± 0.09	99.98% ± 0.05
25 (19 hojas)	7×10^{-7}	100% ± 0	100% ± 0
30 (23 hojas)	5×10^{-7}	100% ± 0	100% ± 0

Table 5. Results from SVM using a HMM with 50 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	5×10^{-7}	99.98% ± 0.01	99.98% ± 0.01
10 (8 hojas)	2×10^{-6}	99.96% ± 0.05	99.98% ± 0.04
15 (12 hojas)	7×10^{-7}	99.,98% ± 0.04	100% ± 0
20 (15 hojas)	7×10^{-7}	100% ± 0	100% ± 0
25 (19 hojas)	7×10^{-7}	99.,98% ± 0.05	100% ± 0
30 (23 hojas)	5×10^{-7}	100% ± 0	100% ± 0

Table 6. Results from SVM using a HMM with 60 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	4×10^{-7}	99,91% ± 0,08	99.96% ± 0.05
10 (8 hojas)	7×10^{-7}	100% ± 0	100% ± 0
15 (12 hojas)	7×10^{-7}	99.98% ± 0.04	99.98% ± 0.04
20 (15 hojas)	7×10^{-7}	100% ± 0	100% ± 0
25 (19 hojas)	7×10^{-7}	100% ± 0	100% ± 0
30 (23 hojas)	7×10^{-7}	100% ± 0	100% ± 0

Table 7. Results from SVM using a HMM with 70 states.

% de training	Value of g (RBF)	Lineal	RBF
5 (4 hojas)	1×10^{-6}	99.93% ± 0.07	99.95% ± 0.08
10 (8 hojas)	7×10^{-7}	100% ± 0	100% ± 0
15 (12 hojas)	7×10^{-7}	100% ± 0	100% ± 0
20 (15 hojas)	7×10^{-7}	99.94% ± 0.14	99.94% ± 0.14
25 (19 hojas)	7×10^{-7}	100% ± 0	100% ± 0
30 (23 hojas)	7×10^{-7}	100% ± 0	100% ± 0

Table 8. Results from SVM using a HMM with 80 states.

From the previous tables can be deduced that for a value of 70 states, already good results are found, for low percentages of samples used to train. Only training with 4 leaves (5% of the database) success rates upper than 99.9% (carrying out the test with 71 remaining leaves) are obtained. Therefore, using 70 states in the HMM, it is a good value for the beginning of the operation of this system.

Also it is observed that the result with the RBF kernel, leaves light improvements in the resulting rates. Besides with RBF, the optimum value of the variable "*g*" is always located environment to 10^{-7}, for which shows its stability versus parameters to classify.

7 Conclusion

A novel and very strong system in the application of arboreal leaves has been created, by means of the use of the kernel of Fisher, leaving from the contours of the leaves, and being classified with a SVM, having success rates upper than 99.9%.

The good results obtained, they cause they foretell the use of this technique in the field of the botany, and therefore, the power to facilitate the task to the biologists.

8 Acknowledges

This work has been supported by the research project A/2258/04, from International Cooperation Spanish Agency (AECI), and by the research project UNI2003/26 from University of Las Palmas de Gran Canaria (Spain).

References:

[1]Pedro L Pérez de Paz, Lourdes Negrín Sosa. Phanerogamarum Monographie Tomus XX, Revisión Taxonómica de Sideris L, 1995.

[2]William T. Stearn. Botanical Latin, History Grammar, Syntax, Terminology and Vocabulary, David & Charles. Network Abbot, 1992.

[3]U. Gernander, Y, Chow, D. M. Keenan. Hands: A Pattern Theoretical Study of Biological Shapes, Springer-Verlang, New York. 1991

[4]Lawrence R. Rabiner, "A tutorial on Hidden Markov models and Selected Applications in Speech Recognition", in Proceedings of the IEEE, vol. 77, no. 2, 1989, pp. 257-286.

[5]Lawrence Rabinier, Biing-Hwang Juang. Fundamentals of Speech Recognition. Prentice Hall, Englewood Cliffs, New Jersey. 1993

[6]Jaakkola, T., Diekhans, M., and Haussler, D. (1998). "A discriminative framework for detecting remote protein homologies". Unpublished, available from http://www.cse.ucsc.edu/research/compbio/research.html (visited on 24 September, 2005)

[7]Christopher J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2, number 2, 1998, pp. 121-167.

[8] Miguel Torres, Juan Carlos Briceño, Miguel A. Ferrer Ballester, Itziar Alonso González. Taxonomía de Hojas de Imágenes Foliares Usando una Red Neuronal MLP. XVI Simposium Nacional de la Unión Científica Internacional de Radio, 2001, pp159-160.

[9]Feng Lu, Evangelos E. Milios. Optimal Spline Fitting to Planar Shape. Elsevier Signal Processing No. 37, 1994, pp 129-140.

[10] Sven Loncaric, A Survey of Shape Analysis Techniques. Pattern Recognition, vol 31 No. 8, 1998, pp 983-1001.

[11] Zhaohui Huang, Fernand Cohen. Affine-Invariant B-Spline Moments for Curve Matching. IEEE Transactions on Image Processing, Vol 5, No. 10. 1996, pp 824-836.

12] Z. Bin; L. Yong and X. Shao-Wei, "Support vector machine and its application in handwritten numeral recognition", *Proceedings of the 15th International Conference on Pattern Recognition*, vol. 2, 2000, pp. 720-723.