TREE SPECIES CLASSIFICATION USING RADIOMETRY, TEXTURE AND SHAPE BASED FEATURES

Maria Kulikova, Meena Mani, Anuj Srivastava, Xavier Descombes, Josiane Zerubia

TREE SPECIES CLASSIFICATION USING RADIOMETRY, TEXTURE AND SHAPE BASED FEATURES

Maria S. Kulikova[1], *Meena Mani*[1], *Anuj Srivastava*[2], *Xavier Descombes*[1], *Josiane Zerubia*[1]

[1] Ariana Research Group, INRIA/I3S
06902 Sophia Antipolis Cedex France
e-mail: firstname.lastname@inria.fr,
mmani@ucla.edu to join Meena Mani

[2] Department of Statistics
Florida State University
Tallahassee, FL 32306
e-mail: anuj@ani.stat.fsu.edu

ABSTRACT

We consider the problem of tree species classification from high resolution aerial images based on radiometry, texture and a shape modeling. We use the notion of shape space proposed by Klassen et al., which provides a shape description invariant to translation, rotation and scaling.

The shape features are extracted within a geodesic distance in the shape space. We then perform a classification using a SVM approach. We are able to show that the shape descriptors improve the classification performance relative to a classifier based on radiometric and textural descriptors alone.

We obtain these results using high resolution Colour InfraRed (CIR) aerial images provided by the Swedish University of Agricultural Sciences. The image viewpoint is close to the nadir, i.e. the tree crowns are seen from above.

1. INTRODUCTION

Interest in applying remote sensing to forests goes back to the 1920s when aerial photographs were first used to assess forest inventory. Remote sensing is now widely employed in forest management where the aerial information is combined with measurements taken on the ground to study the biodiversity of the forest ecosystem. The methods developed in forest image processing aim to facilitate the task of forest inventory and assessment.

The most useful parameters obtained from aerial images and ground measurements are density (of planting), age of trees, stem volume, tree species composition, and information such as biotops and habitats that have ecological value.

In order to obtain information about the diversity of the forest species and stem volume, the classification of tree crowns into species is necessary. Prior to classification, we need to segment the image into individual tree crowns. Segmentation techniques such as template matching, edge detection and others that have been employed for this application are discussed in [5], [13].

A few approaches have been proposed to classify the trees into species. One method is the *Signature Generation Process* where for every crown extracted, a class of signatures is created from the multispectral data of initial image, (cf [10]). A likelihood maximisation technique is used to label the crown. Some crowns with signatures too far from a succesful match remain unclassified. In another study [6], tree crowns are manually delineated to avoid bias due to bad detection. 50 trees for each class verified against the ground truth are selected. 7 parameters such as the multispectral average of pixels in the crown (average on each of the bands) and the illuminated part of the crown, as well as the multispectral value of the tree crown (most lighted pixel) are then computed. Spectral signatures of a crown or a region within a crown are developed by combining means and covariance patterns. The distinct characteristics allow us to regroup delineated crowns in forested populations [7]. There are some limitations with this method due to the close spectral signature of species like the red cedar and the fir for instance [8].

In [4], a classification based on reflectance is used to separate the conifers from the deciduous trees. The internal structure and shading within a crown offer other differentiating criteria. One such measure, the proportion of red and clear pixels to the total number of pixels, can identify birch trees. Another helps differentiate aspens from spruces. A hierarchy of criteria is set forth to classify the crowns. Classification accuracy using this strategy was 75%.

Radiometry and texture analysis have been used extensively in remote sensing applications. In this paper, we incorporate information obtained from studying **tree crown shapes** to improve classification performance.

As in the Erikson [4] study, the tree classification is performed on the four most prevalent species of in Sweden: Norway spruce, Scots pine, birch, and aspen. Two of these species are coniferous while the other two are deciduous. We select 48 crowns (12 per class) from the high resolution CIR images (Fig.1). Their contours, represented by a set of points, are then extracted in order to study their shapes.

The support vector machine (SVM), a supervised learning method, was chosen for the classification. An important property of this classifier is that during the training process, only a small subset of the training set vectors are selected as support vectors. This reduces the computational cost and provides better generalization, so that, for instance, when new samples far from the decision boundary are introduced, the existing support vectors remain unchanged.

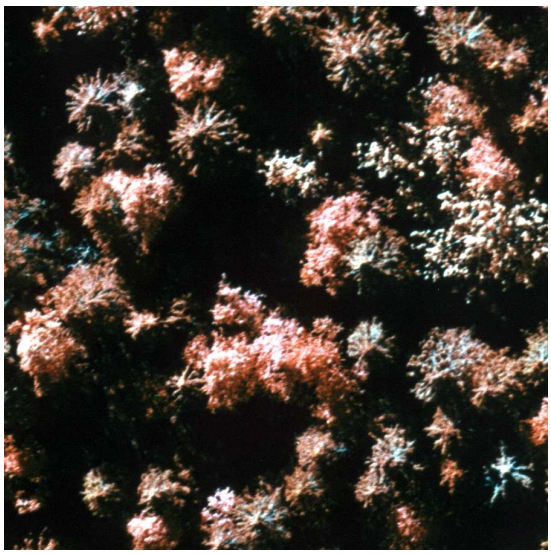

Figure 1: An example of one of the images used in this work (resolution 3 cm), ©Swedish University of Agricultural Sciences.

The work of the first author has been partly supported by a grant of the French Government. This work has been conducted within the INRIA ARC "Mode de Vie" joint research program (http://www-sop.inria.fr/ariana/Projets/ModedeVie/) and INRIA/FSU associated team "SHAPES" (http://www-sop.inria.fr/ariana/Projets/Shapes/).

2. SUPPORT VECTOR MACHINE

SVMs are a set of supervised learning methods of classification and regression, which consists in finding the maximum separation (margin) between classes using a set of observations called the training data. SVMs are also called maximum margin classifiers since they minimize the empirical classification error and maximize the geometric margin simultaneously [15].

2.1 Linear SVM

2.1.1 Separable case

Given a training data (x_i, y_i), $i = 1,...,N$, where $x_i \in R^m$ and $y_i \in \{-1, 1\}$, that denotes to which class x_i belongs, SVM looks for the Optimal Separating Hyperplane that maximizes the distance between the closest training samples of the two classes called also the "margin".

The dividing hyperplane is defined as following $w \cdot x + b = 0$. The vector w is a vector normal to the hyperplane and b is the offset parameter allowing us to increase the margin. So the classifier is given by $f : x \in R^m \longmapsto sign(w \cdot x + b) \in \{-1, 1\}$, i.e. all the training data satisfy the following constraints:

$$\begin{cases} w \cdot x_i + b \geq 0 \text{ if } y_i = +1 \\ w \cdot x_i + b \leq 0 \text{ if } y_i = -1 \end{cases}$$

That could be described by a set of inequalities:

$$y_i(w \cdot x_i + b) \geq 1 \ , \ \forall i \tag{1}$$

To find the hyperplane which gives the maximum margin $2/\|w\|$ (for details see [2]) we should minimize $\|w\|^2$, subject to constraints (1). This leads to the following quadratic optimization problem:

$$\min_{(w,b)} \frac{\|w\|^2}{2}$$

By introducing the Lagrange multipliers $\lambda_i, i = 1,..,N$, one for each inequality constraints (1), the problem becomes "dual" [2]:

$$\max_{\lambda}(L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i \cdot x_j),$$
$$\text{subject to } \sum_{i=1}^{N} \lambda_i y_i = 0 \ , \ 0 \leq \lambda_i \ , \ \forall i,$$

which is a convex quadratic optimization problem subject to linear constraints. Thus, the solution is given by $w = \sum_{i=1}^{N} \lambda_i y_i x_i$ and $b = y_i - w \cdot x_i$, for $i : \lambda_i \neq 0$. The classification function becomes:

$$f(x) = \sum_{i=1}^{N} \lambda_i y_i x_i \cdot x + b$$

2.1.2 Nonseparable case

For nonlinearly separable data, slack variables ξ_i and a regularization parameter C are introduced to deal with misclassified samples, i.e. to relax the constraints (see Fig. 2). By introducing them into the optimization problem, we obtain:

$$\min_{(w,b)} \frac{\|w\|^2}{2} + C \sum_{i=1}^{N} \xi_i,$$

$$\text{subject to : } y_i(w \cdot x_i + b) \geq 1 - \xi_i \ , \ \xi_i \geq 0, \ \forall i.$$

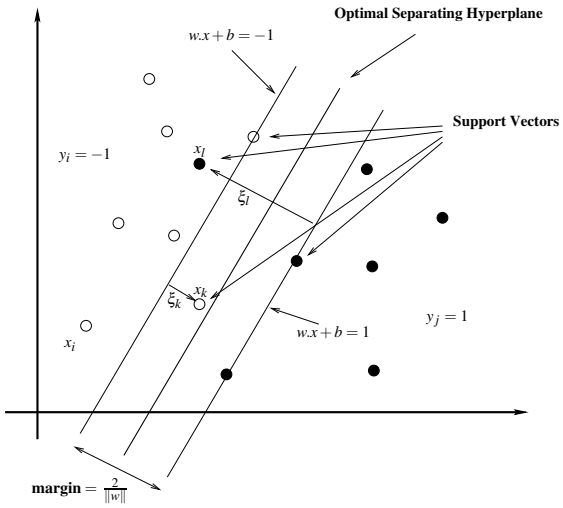

Figure 2: SVM Classifier.

This quadratic problem using the Lagrange multipliers becomes:

$$\max_{\lambda}(L(\lambda) = \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i \cdot x_j),$$
$$\text{subject to } \sum_{i=1}^{N} \lambda_i y_i = 0 \ , \ 0 \leq \lambda_i \leq C \ , \ \forall i.$$

2.2 Nonlinear SVM

For the cases where the decision function is not a linear function of data, the nonlinear classifier was created by applying the kernel trick (originally proposed by Aizerman) to maximum-margin hyperplanes (cf [1]). The resulting algorithm is formally similar, except that every dot product $x_i \cdot x_j$ is replaced by a non-linear kernel function $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$, with $\Phi : R^m \longmapsto H$ and H being an Euclidean space generally of a higher dimension. This allows the algorithm to fit the maximum-margin linear hyperplane in the transformed H.

Thus, the classification function becomes:

$$f(x) = \sum_{i=1}^{N} \lambda_i y_i K(x_i, x) + b.$$

In regard to kernels, there exists a mapping Φ, if $K(x_i, x_j)$ satisfy the Mercer's condition [16].

3. FEATURE CREATION AND CLASSIFICATION RESULTS

As mentioned in the introduction, 48 tree crown contours were selected. The contours were carefully delineated manually to preserve important tree crown shape information. The SVM classification is performed using a Gaussian kernel $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2})$. Since the database is quite small, for each experimental run, 50% of the samples were picked at random to form the training set.

First, we performed the classification using only radiometric attributes. We then add texture features and finally, we included the shape descriptors. After each step, we calculated the performance.

To evaluate the performance, the average performance, P, of a set of experiments is computed. 5% of the values at the high and low end of the performance scale are excluded from this calculation. P can be expressed as

$$P = (\sum_{i=1}^{N_e} P_i - \sum_{w=1}^{N_w} P_w - \sum_{b=1}^{N_b} P_b)/(N_e - N_w - N_b).$$

$P_i = N_c/N$ is the performance of experiment, N_e the number of experiments, N_c the number of correctly classified trees, N the number of trees and N_w and N_b the numbers of the worst and the best results respectively. The maximum performance P_{max} is also given (from which we can get the best combination for training and test sets). Then, we present the *confusion matrix* for the best performance run. A confusion matrix is a visual representation of actual versus predicted classifications.

3.1 Radiometry based features

For vegetation and land use monitoring, CIR or false-colour film offers a richer set of information than natural colour film. The false colours that describe the three channels are Red(NIR), Green(Red), Blue(Green), [12]. CIR photography is commonly used to discriminate live healthy trees from dying vegetation.

In our study, CIR allows us to distinguish the conifers from deciduous trees. As was pointed out in [4], the 4 classes of trees - aspen, birch, spruce and pine - can be easily identified as deciduous or coniferous from first order statistics (the mean and standard deviation, computed from the histogram of pixel intensities on the image). This is because deciduous trees reflect a substantially greater percentage of infrared light.

Classification based exclusively on these descriptors gives low performance with the average performance being $P = 0.54$ and the maximum, $P_{max} = 0.67$. The confusion matrix for one of the best values of P after repeated classification runs is shown below

$$\begin{bmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.167 & 0.666 & 0.167 & 0 \\ 0 & 0.334 & 0.666 & 0 \\ 0 & 0 & 0.167 & 0.833 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$

where (a)aspen, (b)birch, (c)spruce, (d)pine.

From this matrix, we can see that there are a substantial number of false positives. For example, 50% of aspens are classified as birch, while 33% of the birch trees are classified as either aspen or spruce. This classifier is especially weak in differentiating between coniferous and deciduous classes, i.e. between aspen and birch or between spruce and pine.

3.2 Texture based features

To further distinguish within the deciduous and coniferous classes, we performed texture analysis using gray level co-occurrence matrices (GLCM), which is well adapted for characterising microtextures. A co-occurrence matrix is a two-dimensional quantitative representation of spatial relationship [9].

Let $\{I(x,y), 0 \le x \le N-1, 0 \le y \le N-1\}$ denote a $N \times N$ image with G gray levels as described in [3]. The $G \times G$ gray level co-occurrence matrix $P_{\mathbf{d}}$ is defined as

$$P_{\mathbf{d}}(i,j) = |((r,s),(t,v)) : I(r,s) = i, I(t,v) = j|.$$

The entry (i,j) of matrix $P_{\mathbf{d}}$ is the number of occurrences of the pair of gray levels i and j which are a distance $\mathbf{d} = (dx,dy)$ apart. Since $\mathbf{d}$ is a displacement vector, (r,s), $(t,v) \in N \times N$, such that $(t,v) = (r+dx, s+dy)$, and $\|.\|$ is the cardinality of a set. GLCMs are a compact representation of pairs of pixel values in relation to each other. They are an example of the second order statistics as defined by Julesz and several useful features can be computed from them.

We generated 9 such matrices for each tree, each matrix representing a different direction or distance. Two texture features, *energy* and *contrast* were extracted from the GLCMs.
The energy term, given by

$$\sum_i \sum_j P_{\mathbf{d}}^2(i,j)$$

is a measure of the homogeneity of the texture. If the gray level transitions are roughly uniformly distributed, the energy has a low value. Conversely, textures which have dominant gray level transition modes have higher energy values.
The contrast feature

$$\sum_i \sum_j (i-j)^2 P_{\mathbf{d}}(i,j)$$

is a measure of local variation present in an image. We selected this feature to exploit the distinctive features present in the four types of crown surface. Spruce trees, for example, display a radial pattern while aspen have random light and dark regions.

Since these two features are independent of the size and shape of the crown surface, they represent pure texture characteristics.

The evaluation of classifier performance allows us to determine the optimal couple of parameters (d and direction). In practice we have chosen $d = 1$ and 135 degree direction.

By incorporating these two texture features into the classifier, we were able to separate the trees into 4 classes with the average performance $P = 0.71$ and the maximum performance $P_{max} = 0.833$. The confusion matrix for one the best values of P after repeated classification runs is shown below:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.833 & 0 & 0.167 \\ 0.334 & 0 & 0.666 & 0 \\ 0 & 0 & 0.167 & 0.833 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$

We can see that the texture information allows to improve the classification results for desciduous, especially for aspens (a).

3.3 Shape based features

To further improve the previous results, we propose to study tree crown shapes in a so called "shape space", using the representation of planar shapes by their angle functions, cf [11].

We consider the tree crown contours as continuous and closed curves in $\mathbb{R}^2$. The representation of such curves in the shape space is invariant to rigid rotation and translation, and to scaling in $\mathbb{R}^2$. Let us define these properties.

Curves $\alpha = (\alpha_1(s), \alpha_2(s))$ are parametrized by arclength s, where $\alpha : \mathbb{R} \to \mathbb{R}^2$ with period 2π satisfying condition of constant speed along the curve $|\alpha'(s)| = 1$, $\forall s$. We can write $\alpha'(s) = e^{j\theta(s)}$, where $\theta : \mathbb{R} \to \mathbb{R}$ and $j = \sqrt{-1}$ associating $\mathbb{C}$ with $\mathbb{R}^2$. θ is called *direction function* or *angle function*. For every s, $\theta(s)$ gives the angle between $\alpha'(s)$ and the positive abscissa x, cf Fig.3:

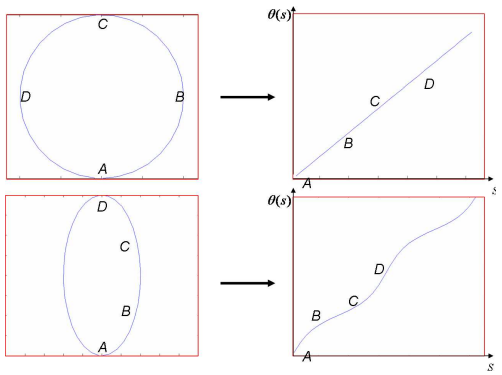

Figure 3: Examples of shapes and corresponding angle functions.

On the Fig.4 we can see more complicated graphs of angle function representing some crowns.

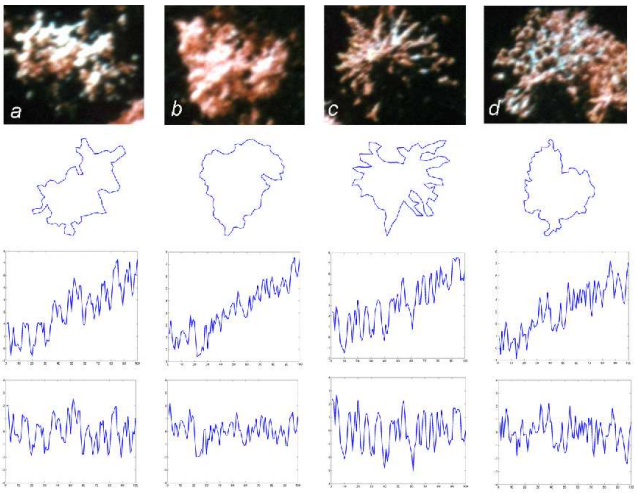

Figure 4: Examples of crowns of four species, their contours, corresponding angle functions θ, and $\tilde{\theta} = \theta - \theta_0$ at the bottom. (a)Aspen, (b)Birch, (c)Spruce, (d)Pine.

We assume $\mathbb{L}^2$ to denote the space of real functions $\mathbb{R} \to \mathbb{R}$ of the period 2π and square integrable on $[0, 2\pi]$ with scalar product $\langle f_1, f_2 \rangle = \int_0^{2\pi} f_1(s) f_2(s) ds$ and $\|f\| = \sqrt{\langle f, f \rangle}$.

Let $\theta(s)$ be the angle function of a planar shape. For the unit circle the function is $\theta_0(s) = s$. For any other closed curve the angle function can be written as $\theta = \theta_0 + f$, where $f \in \mathbb{L}^2$. The space $\theta_0 + \mathbb{L}^2$ is an affine space the elements of which are different by an element of $\mathbb{L}^2$.

The properties of curve invariance mentioned above are guaranteed by the following conditions (cf [11]):

1. The problem of scaling can be simply resolved by fixing the length of the curve, for instance by 2π. Let $\bar{c}$ be a closed contour, then we obtain:
$$\int_{\bar{c}} ds = \int_0^{2\pi} ds = 2\pi;$$

2. An addition of a constant to the angle function θ is equivalent to the rotation of the curve in $\mathbb{R}^2$. To guarantee the invariance of the curve to this action, we deal with the functions θ, the mean values of which are equal to a constant on $[0, 2\pi]$. So, let $\frac{1}{2\pi} \int_0^{2\pi} \theta(s) ds = \pi$, where constant π is chosen to include the identity function θ_0 in the restricted set, since
$$\frac{1}{2\pi} \int_0^{2\pi} \theta_0(s) ds = \frac{1}{2\pi} \int_0^{2\pi} s ds = \pi;$$

3. Finally, to be closed, the curves must satisfy following condition:
$$\int_0^{2\pi} \exp(j\theta(s)) ds = 0.$$

We define $C \subset \theta_0 + \mathbb{L}^2$ as the set of all the elements of $\theta_0 + \mathbb{L}^2$ satisfying the conditions 1-3 described above. Or more formally, define the map $\phi = (\phi^1, \phi^2, \phi^3)) : (\theta_0 + \mathbb{L}^2) \to \mathbb{R}^3$:

$$\phi^1 = \frac{1}{2\pi} \int_0^{2\pi} \theta(s) ds$$

$$\phi^2 = \frac{1}{2\pi} \int_0^{2\pi} cos(\theta(s)) ds$$

$$\phi^3 = \frac{1}{2\pi} \int_0^{2\pi} sin(\theta(s)) ds.$$

From now on, we can define C as $C = \phi^{-1}(\pi, 0, 0)$, which is called the pre-shape space, because the same curve with different initial points ($s = 0$) correspond to different elements of the space C. But we will deal with the shape space S, where such curves are represented by one element of S defined by C/D, where D is the unite circle.

We are going to describe the tree features created using such a representation of the crown contours. To do it some tree crown shape analysis was done to determine the information that could allow us to recognize the species.

Let us look at Fig.4. We can see that the aspens have an irregular structure, the convexities/branches sticking out of the body of crowns. The branches of spruces are more regular and have a more radial direction. The birch and pine crown contours are more round, where birches are the smoothest.

The first feature, chosen to be incorporated in the classifier, is a "distance to a circle" $d(\theta, \theta_0)$, which is computed using the geodesic on the shape space (see [11] for details). For the given shapes $\theta_i, i = 1, ..., M$, we perform a vector of "distances to a circle":
$$v_c(\theta) = \{d(\theta_i, \theta_0), \ \ i = 1, 2, \ldots, M\} \in \mathbb{R},$$
where M is the number of trees.

We also include geometrical descriptors based on $\tilde{\theta} = \theta - \theta_0$.

Firstly, we translate the property that some species have more regular structure of the crowns than others, by a measure of contour elasticity:
$$v_e(\tilde{\theta}) = \int_0^{2\pi} \dot{\tilde{\theta}}(s)^2 ds,$$
$$v_e(\tilde{\theta}) = \{v_e(\tilde{\theta}_i), \ \ i = 1, 2, \ldots, M\} \in \mathbb{R}.$$

In regard to spruces, generally they have big branches/convexities and not numerous in comparison with the convexities of birch crown. This criterion is reflected by the angle function under the form of the local maxima number N.
$$v_N(\tilde{\theta}) = N,$$
$$v_N(\tilde{\theta}) = \{v_N(\tilde{\theta}_i), \ \ i = 1, 2, \ldots, M\} \in \mathbb{R}.$$

Then, the pine crown shape is quite close to that of birch but certain roughnesses are a little bit bigger, some branches sticking out. Thus we can qualify the crown contour irregularities due to the branches, leaves and shadows:
$$v_\mu(\tilde{\theta}) = \mu = \frac{1}{n} \sum_{k=1}^{n} \left| \tilde{\theta}(s_k) \right|,$$
$$v_\mu(\tilde{\theta}) = \{\mu_i, \ \ i = 1, 2, \ldots, M\} \in \mathbb{R},$$
where $s_k = \sum_{t=1}^{k} c_t$, $c_t = \|p_{t+1} - p_t\|$ chord lengths of contour and $p_t \in \mathbb{R}^2, t = 1, ..., n$ ordered collection of points approximating the contours,
$$v_{Var}(\tilde{\theta}) = Var = \frac{1}{n-1} \sum_{k=1}^{n} (\mu_i - \left| \tilde{\theta}(s_k) \right|)^2,$$
$$v_{Var}(\tilde{\theta}) = \{Var_i, \ \ i = 1, 2, \ldots, M\} \in \mathbb{R}.$$

Finally, we use the density function of the angle function in order to have information about the quantity of big and small convexities of the crown contour. For that, we calculate the histograms of $\tilde{\theta}$, $h_{\tilde{\theta}}(x) = card\{\tilde{\theta}(s) = x, s \in [0, 2\pi], x \in \mathbb{R}\}$, and then Euclidean distances
$$d(h_1(x), h_2(x)) = \sqrt{< h_1(x), h_2(x) >}.$$
Then we create the following features associated to each shape, one distance for every class l. Such a distance is calculated as the distance to the nearest element of the class l:
$$d^l_{min}(h_i) = min\{d(h_i, h_j)\}^{j=1,\ldots m_l} \in \mathbb{R},$$

$$v_d(\tilde{\theta}) = \{d^l_{min}(h_i),\ \ i = 1,2,\ldots,M,\ l = 1,\ldots,N_l\} \in \mathbb{R}\,,$$

where N_l is the number of classes and m_l is the number of shapes of each class in the training set.

Now we are forming a long vector using individual features and apply SVM on that feature vector. Obtained results: the average performance $P = 0.747$ and the maximum of performances $P_{max} = 0.87$. An example of confusion matrix is given as follows:

$$\begin{bmatrix} 0.833 & 0.167 & 0 & 0 \\ 0.167 & 0.833 & 0 & 0 \\ 0 & 0.167 & 0.833 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \end{matrix}$$

Shape based features allow us to improve the identification of species within the conifers and deciduous trees.

4. CONCLUSION

In this paper we show that the performance of a classifier based on conventional spectral and texture characteristics can be improved by including shape descriptors in the feature set. By incorporating shape features, the classification performance mean improved by about 4% for 6 samples per class while the maximum performance was 87.5%.

To create the new descriptors, the shapes of crowns were analysed using the angle function representation. Such representation allowed us to preserve the tree crown characteristics that associate it with one class or another.

The limits imposed by the database size did not give us much freedom to experiment with the training to test sample size ratio. We did, however, observe that the performance mean tended to increase when samples were added to the training set (see fig. 5).

These results are encouraging but they need to be validated on a larger number of images and tree crowns.

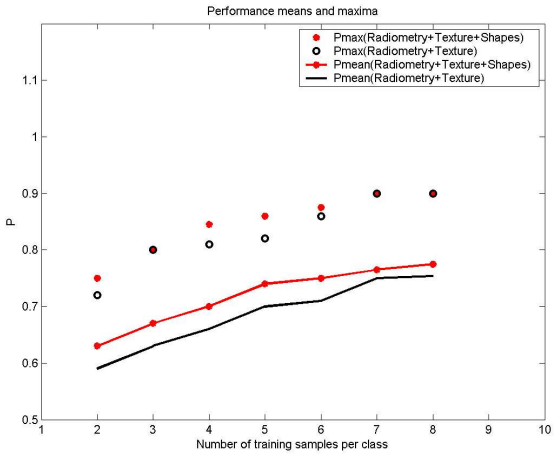

Figure 5: Performance means and maxima.

Future work will consist in automatic extraction of tree crowns using shape information.

REFERENCES

[1] B. E. Boser, I. M. Guyon and V. N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," in *Fifth Annual Workshop on Computational Learning Theory*, Pittsburg, 1992.

[2] C. J. C. Burges, *Tutorial on Support Vector Machines for Pattern recognition*. Kluwer Academic Publishers, Boston. pp. 1–43.

[3] C. H. Chen, L. F. Pau, P. S. P. Wang, The Handbook of Pattern Recognition and Computer Vision (2nd Edition), pp. 207-248, World Scientific Publishing Co., 1998.

[4] M. Erikson, "Segmentation and Classification of Individual Tree Crowns," *Ph.D. thesis*, Swedish University of Agricultural Sciences, Uppsala, Sweden, 2004.

[5] M. Erikson, "Species Classification of Individually Segmented Tree Crowns in High-Resolution Aerial Images using Radiometric and Morphologic Image Measures," *Remote Sensing of Environment*, vol. 91, pp. 469–477, April 2004.

[6] F. A. Gougeon, "Comparison of Possible Multispectral Classification Scheme for Tree Crown Individually Delineated on High Spatial Resolution MEIS Images," *Canadian Journal of Remote Sensing*, 21(1), pp. 1–9, 1995.

[7] F. A. Gougeon, "Vers l'inventaire forestier automatisé : reconnaître l'arbre ou la forêt?," *CD-ROM du 9eme Congrès de L'Association québécoise de télédétection*, May 1996.

[8] F. A. Gougeon, D. G. Leckie, I. Scott and D.Paaradine"Individual Tree Crown Species Recognition: the Nahmint Study," *Proc. of the International Forum on Automated Interpretation of High Spatial Resolution Digital Imagery for Forestry*, pp 209–223, Pacific Forestry Center, Victoria, British Columbia, Canada, February 1998.

[9] R. M. Haralick, "Statistical and Structural Approaches to Texture," *Proceedings of the IEEE*, vol. 67, pp. 786–804, 1978.

[10] D. G. Leckie, F. A. Gougeon, N. Walsworth and D. Paradine, "Stand delineation and Composition estimation using Semi-Automated Individual Tree Crown Analysis," *Remote Sensing of Environment*, vol. 85, pp. 355–369, 2003.

[11] E. Klassen, A. Srivastava, W. Mio and S. H. Joshi, " Analysis of Planar Shapes Using Geodesic Paths on Shape Space," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 26, no 3, pp. 372–383, March 2004.

[12] G. Perrin, "Etude du Couvert Forestier par Processus Ponctuels Marqués," *Ph.D. thesis*, Ecole Doctoral de Centrale Paris, October 2006.

[13] G. Perrin, X. Descombes and J. Zerubia. "A Non-Bayesian Model for Tree Crown Extraction using Marked Point Processes," *Research Report 5846*, INRIA, France, February, 2006.

[14] A. Srivastava, S. H. Joshi, W. Mio and X. Liu, "Statistical Shape Analysis: Clustering, Learning, and Testing," *IEEE Transaction on Pattern Analysis and Machine Intellegence*, vol. 27, no 4, pp. 590–600, March 2004.

[15] V. N. Vapnik, *Statistical Learning Theory*. John Wily and sons, inc., 1998.

[16] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer Verlag, New-York, 1995.