A comparative study of several supervised classifiers for coconut palm trees fields' type mapping on 80 cm RGB pansharpened Ikonos images

Raimana Teina, Dominique Béréziat, Benoît Stoll, Sébastien Chabrier

A comparative study of several supervised classifiers for coconut palm trees fields' type mapping on 80 cm RGB pansharpened Ikonos images

Teina R.[a] and Béréziat D.[a] and Stoll B.[b] and Chabrier S.[b]

[a]Université Pierre et Marie Curie, LIP6, 104 avenue du Président Kennedy, 75016 Paris, France;
[b]Université de la Polynésie Française, Laboratoire GePaSud, 98702 Faa'a, Tahiti - French Polynesia;

ABSTRACT

The purpose of this study is to classify the types of coconut plantation. To this end, we compare several classifiers such as Maximum Likelihood, Minimum Distance, Parallelepiped, Mahalanobis and Support Vector Machines (SVM). The contribution of textural informations and spectral informations increases the separability of different classes and then increases the performance of classification algorithms. Before comparing these algorithms, the optimal windows size, on which the textural information are computed, as well as the SVM parameters are first estimated. Following this study, we conclude that SVM gives very satisfactory results for coconut field type mapping.

Keywords: SVM, Maximum Likelihood, Mahalanobis, Parallelepiped, Minimum Distance, Texture, Principal Component Analysis

1. INTRODUCTION

French Polynesia government wants to improve the coconut tree field exploitation in order to develop the extraction of the Coprah oil as an alternative fuel and also the use of senile trees wood. The French Polynesia Government acquired in 2003 some RGB pansharpened Ikonos images 80 cm pixel resolution on ground for almost the 80 islands in the Tuamotu's archipelago. The goal of this work is to classify the coconut palm trees fields using several classification algorithms and to measure their accuracies. The following classifiers are studied: Maximum Likelihood, Minimum Distance, Mahalanobis Distance, Parallelepiped and Support Vector Machines (SVM). First, a classification based only on spectral informations is not sufficient to achieve the task. Spatial informations also must be considered. Then, Haralick's texture descriptors are computed using the first component of the Principal Component Analysis applied to the spectral bands. These texture descriptors are computed using a window size estimated by analyzing different field types. The measure of classifiers' accuracies is done by analysis of the confusion matrix using a ground truth ROIs. According to this performance analysis, the use of SVM provides a best land use mapping of coconut field types.

2. DATA AND STUDY SITE

The study focuses on the atoll of Vahitahi, which is located in the north-east of the French Polynesia Territory, especially on the North-East area where different types of plantation are available. The remote sensed image was acquired in September 2005 at 19:46 GMT by the satellite Ikonos and projected in the WGS-84 coordinates system. The complete mosaic (10673×4120 pixels) is centered on geographic coordinates $138°51'52.96''W$ and $18°45'35.19''S$. The RGB images are pansharpened with a ground resolution of the pixel of 80 cm. Figure 1 illustrates a global view of the atoll of Vahitahi and a zoom of the study area in the North-East.

Further author information: (Send correspondence to Teina R.)
Teina R.: E-mail: Raimana.Teina@lip6.fr, Telephone: +33 (0) 1 44 27 88 16
Béréziat D.: E-mail: Dominique.Bereziat@lip6.fr

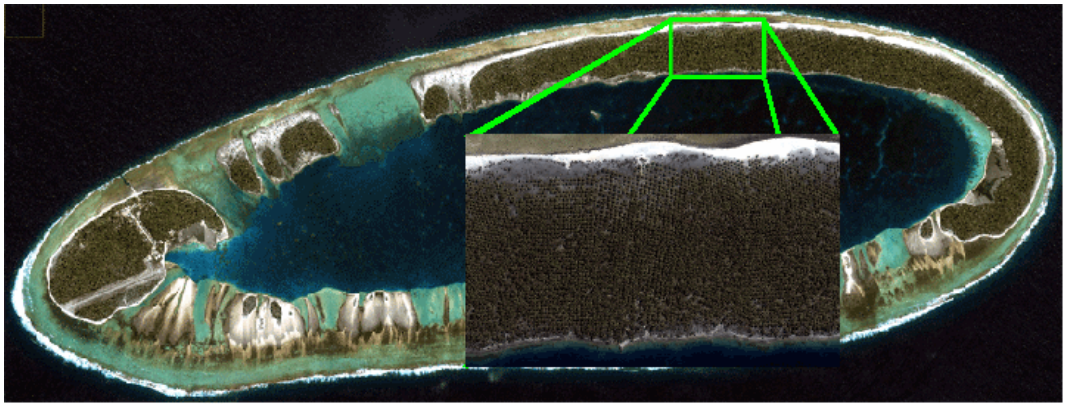


Figure 1. The atoll of Vahitahi and a zoom of the study site on the north-east.

3. CLASSIFIERS REVIEW

In this section, we review the most common supervised classifiers used in remote sensing images analysis. Because of the complexity of the structures to be segmented, unsupervised classifiers is not appropriated in our experimental context and we need to build training set for each class of field we want to classify.

3.1 Maximum Likekihood Classification

Let the M classes be represented by $\omega_i,\ i = 1, ..., M$. The conditional probability for a pixel at location x belongs to class ω_i is $p(\omega_i|x)$. The classification is performed according to:

$$x \in \omega_i, \text{if } p(\omega_i|x) > p(\omega_j|x) \text{ for all } j \neq i \tag{1}$$

The problem is now to compute the conditional probabilities. Using Bayes' rule, we have:

$$p(\omega_i|x) = p(x|\omega_i)p(\omega_i)/p(x) \tag{2}$$

$p(x|\omega_i)$ is estimated from the training set and using Equation (2), the Maximum Likelihood (ML) classification's rule defined in Equation (1) becomes:

$$x \in \omega_i, \text{if } p(x|\omega_i)p(\omega_i) > p(x|\omega_j)p(\omega_j) \text{ for all } j \neq i \tag{3}$$

For mathematical convenience results, a discriminant function g_i is defined:

$$\begin{aligned} g_i(x) &= \log\{p(x|\omega_i)p(\omega_i)\} \\ &= \log p(x|\omega_i) + \log p(\omega_i) \end{aligned} \tag{4}$$

Equation (3) becomes:

$$x \in \omega_i, \text{if } g_i(x) > g_j(x) \text{ for all } j \neq i \tag{5}$$

We assume a Gaussian distribution for $p(x|\omega_i)$:

$$p(x|\omega_i) = (2\pi)^{-N/2}|\Sigma_i|^{-1/2} \exp\{-\frac{1}{2}(x - m_i)^t \Sigma_i^{-1}(x - m_i)\} \tag{6}$$

where m_i and Σ_i are the mean vector and covariance matrix of the data in class ω_i. The discriminant function is then:

$$g_i(x) = \log p(\omega_i) - \frac{1}{2}\log|\Sigma_i| - \frac{1}{2}(x - m_i)^t \Sigma_i^{-1}(x - m_i) \tag{7}$$

Assuming an equal probabilities distribution ω_i, term $\log p(\omega_i)$ can be removed from Equation (7) as it is without effect on the decision rule:

$$g_i(x) = -\log|\Sigma_i| - (x - m_i)^t \Sigma_i^{-1}(x - m_i) \tag{8}$$

3.2 Minimum Distance and Mahalanobis Classification

Let $m_i, i = 1, ..., M$ be the means of the M classes determined from the training set, and let x be the location of the pixel to be classified. The Euclidean distance of x to each class is computed: $d(x, m_i)^2 = (x - m_i)^T(x - m_i)$. The Minimum Distance (MinDist) classification is performed on the base of the following rule:

$$x \in \omega_i, \text{if } d(x, m_i)^2 < d(x, m_j)^2 \text{ for all } j \neq i \tag{9}$$

An alternative to Euclidean distance can be considered using the discriminant function defined in Equation (8):

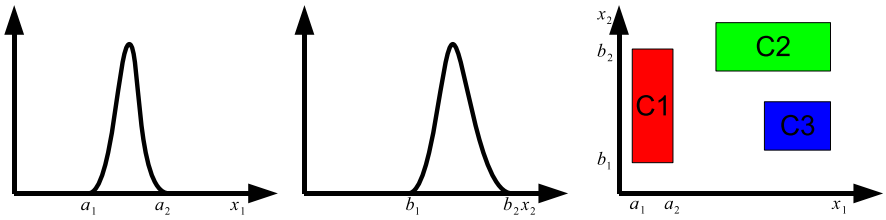
$$d(x, m_i)^2 = \ln |\Sigma_i| + (x - m_i)^t \Sigma_i^{-1} (x - m_i) \tag{10}$$

In the cases where all class covariances are equal, i.e $\Sigma_i = \Sigma$ for all i, the term $\log |\Sigma|$ can be ignored. The Mahalanobis (Mahal) distance classifier is defined using the following distance measure:

$$d(x, m_i)^2 = (x - m_i)^t \Sigma^{-1} (x - m_i) \tag{11}$$

3.3 Parallelepiped Classification

The parallelepiped (Para) classifier analyzes histograms of the individual spectral component. The lower and the upper significant bounds on the histogram are identified and used to describe the brightness value range for each band for that class. The range in all bands describes a multidimensional box or parallelepiped. These boxes are obtained from the training set. During the classification process, if pixels are found to lie in such a parallelepiped they are labelled as belonging to that class. This classification algorithm is simple to train and to use, it however has several drawbacks. First is that there can be gaps between the boxes and pixels lying in those regions couldn't be classified. Second, correlated data can be overlap of the boxes, pixels in those regions can not be separated and are unclassified.



(a) Histograms with upper and lower bounds.

(b) Examples of 2D parallelepipeds.

Figure 2. Principle of the Parallelepiped classification.

3.4 Support Vector Machine (SVM)

The SVM classifier was introduced in 1995 by Vapnik.[1] This Subsection briefly describes this classifier, details can be found in[2] and.[3] Figure 3 illustrates the principle of the optimal hyperplan and the optimal margin used in the SVM classifier.

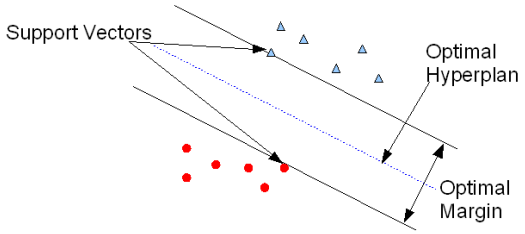


Figure 3. SVM Classifier principle.

Let's consider the case of a two-label classification problem. A training set data is constituted of N samples described by Support Vectors (SV) x_i and associated labels y_i. A label takes -1 or $+1$ as values. The SVM classifier consists in defining the function:

$$f(x) = \langle \omega, x \rangle + b \tag{12}$$

which maximizes the margin between the optimal hyperplan and the support vectors.
The hyperplan equation is $\langle \omega, x \rangle + b = 0$, where ω is the normal to the hyperplan and $\frac{|b|}{||\omega||}$ is the perpendicular distance from the SV to the optimal hyperplan. The sign of $f(x)$ gives the label of the sample. The goal of the SVM is to maximize the margin between the optimal hyperplan and the support vectors, i.e the research of $min \frac{||\omega||}{2}$. The problem can be solved by using the Lagrange multipliers:

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} y_i.\alpha_i \langle x, x_i \rangle + b\right) \tag{13}$$

where α_i are Lagrange multipliers.

In the case of nonlinear separable data, one possible method to solve the problem is to use a kernel. A kernel is a function which project data to be classified into a higher dimension space feature, $\Phi : \mathbb{R}^n \rightarrow \mathbb{F}$, in a such way that data are now considered as linearly separable. In Equation (13), the dot product $\langle x, x_i \rangle$ is replaced by the dot product associated to the space feature $\mathbb{F}$ defined as:

$$K(x, x_i) = \langle \Phi(x), \Phi(x_i) \rangle \tag{14}$$

and then the function to classify data becomes

$$f(x) = \text{sign}\left(\sum_{i=1}^{N} y_i.\alpha_i.K(x, x_i) + b\right) \tag{15}$$

In practise, the kernel function is difficult to define but the dot product K can be defined more easily. In this study, we use the Radial Basis Kernel (RBF) defined with $K(x, x_i) = e^{-\frac{|x-x_i|^2}{2\sigma^2}}$. This kernel gives the best results in our case compared to other kernels we have tested.

4. METHODOLOGY

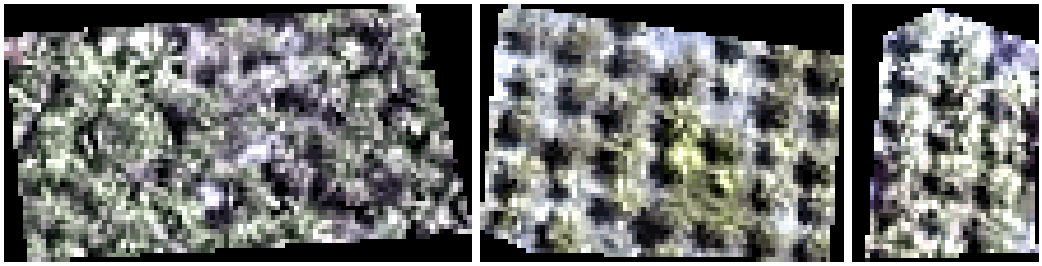
4.1 Training sets

The goal of the study is to analyze the accuracy of the supervised classifiers for the characterization of different planting types. We first segment the vegetation part where we define regions of interest (ROI) corresponding to each class of fields. As recommended by Swain and Davis,[4] $100 \times N$ (with N the number of bands) training pixels per class are needed to avoid the covariance matrix to be singular. In our study, training sets are built by randomly generating 1500 pixels in each ROI. Figure 4 shows three classes of field types listed below:

- 4(a): natural field where there is not any spacial organization between trees (Natural),
- 4(b): artificial field with low density (Artificial LD),
- 4(c): artificial field with medium density (Artificial MD).

To predict the result of the final classification, we compute the classes separability with the Jeffries-Matusita (JM) distance. This measure is the average distance between two classes density function (Wacker[5]) and for normally distributed classes it is defined as:

$$J_{ij} = 2(1 - e^{-B}) \tag{16}$$



(a) Natural (b) Artificial LD (c) Artificial MD

Figure 4. Training set for each class of plantation.

where B is referred to the Bhattacharyya distance (Kailath[6]) and is defined as

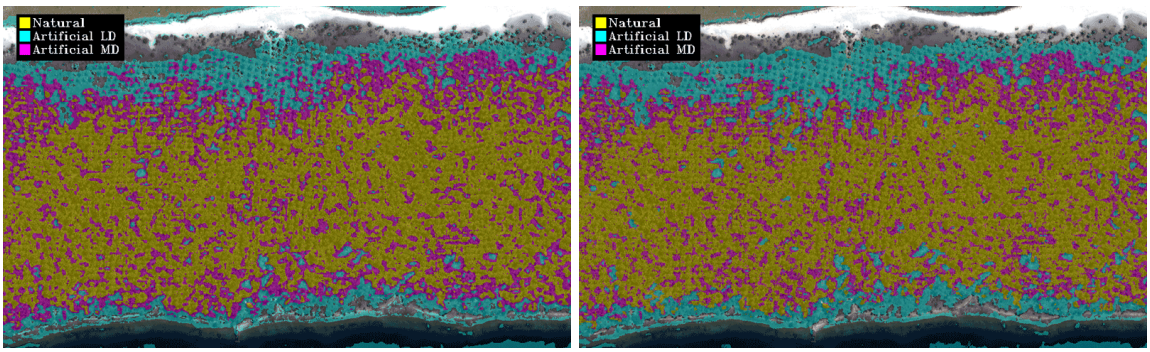
$$B = \frac{1}{8}(m_i - m_j)^t \left(\frac{\Sigma_i + \Sigma_j}{2}\right)^{-1} (m_i - m_j) + \frac{1}{2}\ln\left(\frac{|(\Sigma_i + \Sigma_j)|}{2\sqrt{\Sigma_i \Sigma_j}}\right) \tag{17}$$

As shown in Tab. 1, the use of only spectral informations is not enough to achieve a good classification. We can see that the best pair separation doesn't exceed 1.446 and at least Artificial LD and Artificial MD only have a pair separation equal to 0.407.

Table 1. Jeffrey-Matusita distance for each class.

	Natural	**Artificial LD**	**Artificial MD**
Natural	0		
Artificial LD	1.44641	0	
Artificial MD	0.74986	0.40689	0

The spectral information is then not sufficient to characterize the fields types. Indeed Figure 5 below, where classifications have been made only on spectral informations with ML and SVM algorithms, shows unsatisfactory results.



(a) ML classification (b) SVM classification

Figure 5. ML and SVM classification's results using only the spectral informations.

In addition of spectral informations, we propose to use the spatial information contained in the image. This spatial information can be obtained from Haralick's texture descriptors (taking into account spatial and spectral informations).

4.2 Textural informations

A Principal Component Analysis (PCA) transforms a multichannel information into another multichannel representation in which, each channel being uncorrelated with other channels. Ricotta et al.[7] showed that the first

component of PCA representation contains exhaustive informations required to perform the image analysis. The texture descriptors are then computed on this component.

The eight texture descriptors described by Haralick[8–11](GLCM) and used in this work are: mean, variance, homogeneity, contrast, dissimilarity, entropy, second moment and correlation. As the artificial plantation have no preferred direction, each texture descriptor is computed in four directions ($0°, 45°, 90°$ and $135°$) than a mean descriptor is calculated by averaging the four directions. The textural informations and the spectral informations are stacked together and normalized.

Figure 6 shows the evolution of the global mean of JM distance versus the size of the window used for the calculation of the texture descriptors. As shown in this figure, the global mean class separability is greater than 1.9 for a window size greater or equal to 29 pixels. This shows that we could obtain a better accuracy of the classification using a window size of 29 pixels.

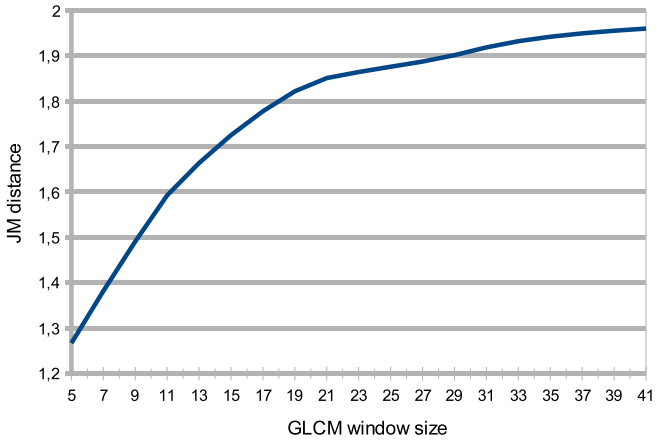


Figure 6. JM distance versus the window size used for the calculation of the texture description.

5. APPLICATION ON IKONOS IMAGES

In this section, we apply the various classifiers reviewed in Sect. 3 using even spectral and textural informations.

5.1 Comparison of different classifiers

For all those methods, we apply the classification with different window sizes (from 5×5 to 41×41 with a step of 2). The accuracy is computed using the coherence matrix computed on 3000 randomly generated pixel's locations in ground truth ROI for each class. The classification's accuracy is then compared by analyzing the mean accuracy, computed by averaging of the diagonal values of the coherence matrix. Tab. 2 shows sample calculation of the mean accuracy (in %).

Table 2. Method to compute the mean accuracy using the confusion matrix.

	Ground Truth			
Classified	Natural	Artificial LD	Artificial MD	
Natural	**98,63**	0	14,33	
Artificial LD	0	**99,89**	16,53	
Artificial MD	1,37	0,11	**69,13**	Mean accuracy
				89,22

Figure 7 shows a plot of the classifiers' accuracies versus the GLCM's window size. As we can see, SVM's and Maximum Likelihood's accuracies are clearly better than the other classifiers. From a GLCM's window size equal 5 to 29, SVM and ML have similar accuracies. But while the GLCM's window size increases, SVM gives some better classification's results than the ML.

The three other classifiers, have lower performance. The Mahalanobis classification's accuracy tends to increase and reaches its maximum for a GLCM's window size of 37 pixels with a mean accuracy value of 74.54%. It shows the influence of textural informations for the classification's performance with a maximum difference between values equal to 24%.
As regards the Parallelepiped algorithm, the curve tends to grow until the GLCM's window size equal 37 pixels where the maximum accuracy is reached with a value of 63.2%.
The Minimum Distance algorithm is one that gives the worst performance. Indeed, we see that the curve has a low slope, although it is positive. And textural informations do not improve more than 7% the accuracy of this method.

Looking at the results, we can distinguish three groups:

- Best performance: SVM and ML,
- Medium performance: Mahalanobis,
- Poor performance: Minimum Distance and Parallelepiped.

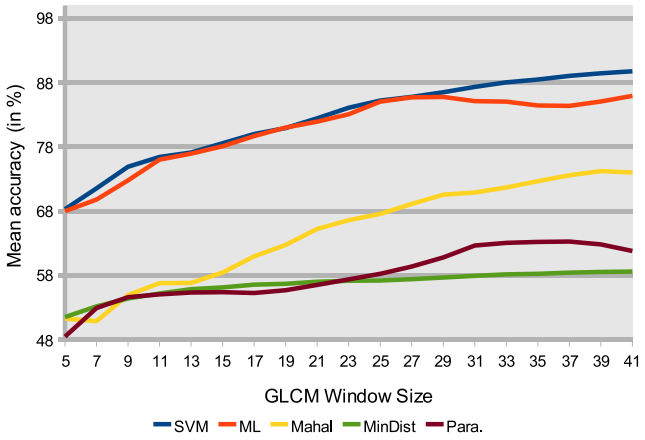


Figure 7. Classifiers' accuracy.

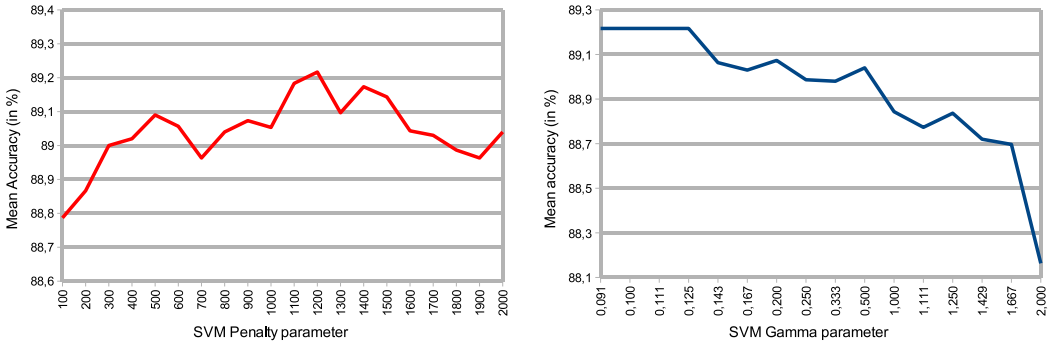
The SVM classification is done using the LIBSVM[12] with the default values, $\gamma = \frac{1}{N}$ with N the number of bands (3 spectral bands and 8 texture descriptors), and penalty set to 100. In the next Subsection, a study of method robustness with respect to these parameters is lead.

5.2 SVM parameters optimization

To highlight the contribution of the SVM's parameters to the accuracy of the classification, we fix a GLCM's window size to 37. The data to classify are still the textural informations and the spectral informations stacked together.

First, the penalty parameter (P) is increased from 100 to 2000 with a step of 100. In Fig. 8(a), the mean accuracy is plotted for each penalty parameter value. Variations of P does not show any significant change in performance. Indeed, accuracy varies between 88.79% and 89.22%. We observe less than 1% of variation. The curve has no constant growth or decay but fluctuates depending on the value of P with a maximum level reached for $P = 1200$ giving 89.22% of mean accuracy.

Next, the penalty parameter is set to 1200 and the γ parameter varies in the range of 100 and 2000 with a step of 0.1. As shown in Fig. 8(b), for γ in range 0.091 and 0.125, we observe constant accuracies. Then, the curve has a tendency to decrease with γ increasing. This shows the important influence of γ on the classification performance.



(a) Penalty (b) Gamma

Figure 8. SVM parameters optimization.

With the help of the previous study of the influence of different parameters, we set the penalty parameter to 1200 and γ to 0091 for our final classification. Fig. 9 illustrates the result of the classification on the study area in the north-east of the atoll of Vahitahi.

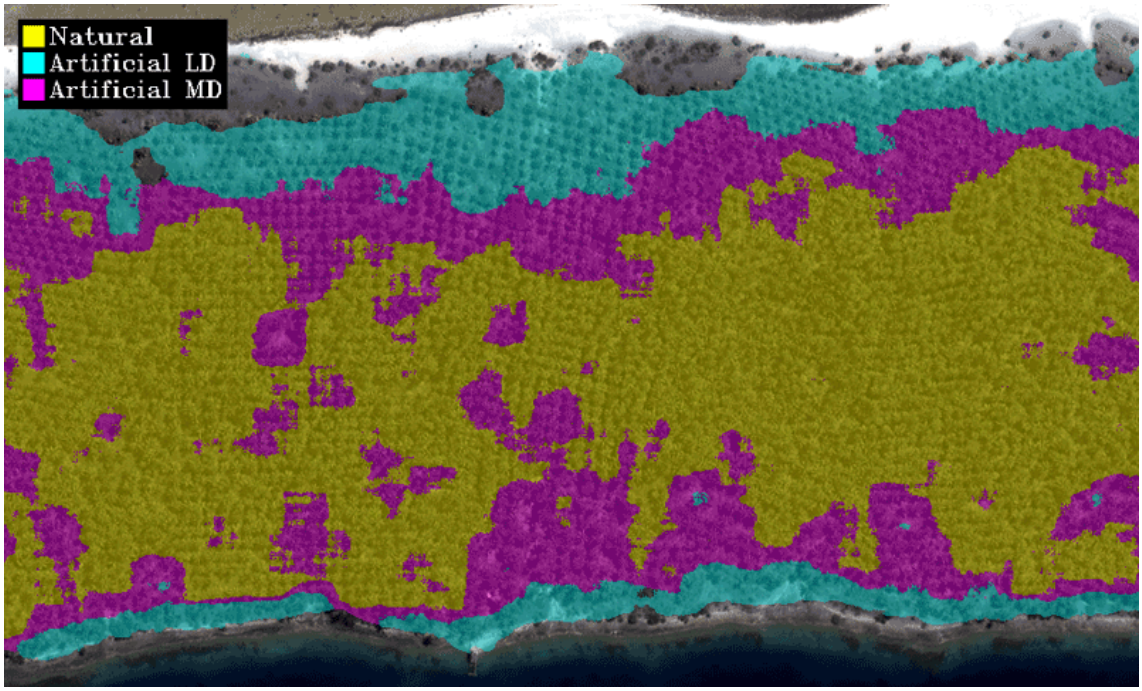


Figure 9. SVM classification using P=1200 and $\gamma = 0.091$.

6. CONCLUSION

In this study, we showed that the use of spectral information not sufficient to classify the different field' types. But the goal is fully reached by providing textural informations. Thanks to the study of classifiers' accuracies, we showed that the use of Minimun Distance and Parallelepiped algorithms produce poor results. Mahalanobis algorithm gives results that are satisfactory but not sufficient to classify correctly fields' types of coconut trees. The best results are obtained with Support Vector Machines and Maximum Likelihood classifications. But, through the study of the influence of various parameters of SVM and although Maximum Likelihood's accuracies are almost equivalent to that of SVM, we showed that SVM gives very satisfactory results with a good selection of parameters P and γ.

The use of the SVM classification is a good tool to estimate from remote sensed images different fields type of coconut trees, and allows an estimation of the number of trees in each class according recommendations about the space between trees in artificial fields.

Because of the lack of the Near Infra-red channel, future works could focus on finding other descriptors based on vegetation indices calculated using the RGB channels or other color space descriptions to improve the quality of classification.

REFERENCES

[1] Cortes, C. and Vapnik, V., "Support-vector networks," *Machine Learning* **20**(3), 273–297 (1995).

[2] Burges, C. J. C., "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery* **2**(2), 121–167 (1998).

[3] Wu, T.-F., Lin, C.-J., and Weng, R. C., "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research* **5**, 975–1005 (2004).

[4] Swain, P. and Davis, S., [*Remote Sensing, the Quantitative Approach*], Mc Graw-Hill, NY (1978).

[5] Wacker, A., *The Minimum Distance Approach to Classification*, PhD thesis, Purdue University, West Lafayette (1971).

[6] Kailath, T., "The divergence and bhattacharyya distance measures in signal selection," *Communications, IEEE Trans. Communication Theory* **15**(1), 52–60 (1967).

[7] Ricotta, C. and Avena, C., "The influence of principal component analysis on the spatial structure of a multispectral dataset," *International Journal of Remote Sensing* **20**(17), 3367–3376 (1999).

[8] Haralick, R. M., Dinstein, and Shanmugam, K., "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-3**, 610–621 (November 1973).

[9] Haralick, R., "Statistical and structural approaches to texture," *Proceedings of the IEEE* **67**, 786–804 (1979).

[10] Anys, H., Bannari, H., He, D., and Morin, D., "Texture analysis for the mapping of urban areas using airborne MEIS-II images," *Proceedings of the First International Airborne Remote Sensing Conference and Exhibition* (3), 231–245 (1994).

[11] Puissant, A., Hirsch, J., and Weber, C., "The utility of texture analysis to improve per-pixel classification for high to very high spatial resolution imagery," *International Journal of Remote Sensing* **26**, 733–745 (February 2005).

[12] Chang, C.-C. and Lin, C.-J., *LIBSVM: a library for support vector machines* (2001). Software avaiable at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.