Support Vector Machines for burnt area discrimination

Olivier Zammit, Xavier Descombes, Josiane Zerubia

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Support Vector Machines for burnt area discrimination

Olivier Zammit — Xavier Descombes — Josiane Zerubia

N° 6343

Novembre 2007

Thème COG

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--6343--FR+ENG

Support Vector Machines for burnt area discrimination

Olivier Zammit , Xavier Descombes , Josiane Zerubia

Thème COG — Systèmes cognitifs
Projet Ariana

Rapport de recherche n° 6343 — Novembre 2007 — 37 pages

Abstract: This report addresses the problem of burnt area discrimination using remote sensing images. The detection is based on a single post-fire image acquired by SPOT 5 satellite. To delineate the burnt areas, we use a recent classification method called Support Vectors Machines (SVM). This approach is compared to more conventional classifiers such as K-means or K-nearest neighbours which are widely used in image processing. We also proposed a new automatic classification approach combining K-means and SVM. The results given by the different methods are finally compared to ground truths on various burnt areas.

Key-words: forest fire, burnt areas, remote sensing, high resolution satellite images, Support Vectors Machines (SVM), K-means, K-nearest neighbours, classification.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France)
Téléphone : +33 4 92 38 77 77 —Télécopie : +33 4 92 38 77 65

Séparateurs à Vaste Marge pour la discrimination des zones brûlées

Résumé : Ce rapport aborde le problème de l'évaluation des dégâts après un feux de forêt. La détection est effectuée à partir d'une seule image satellite (SPOT 5) acquise après le feu. Afin de détecter les zones brûlées, nous utilisons une approche récente de classification nommée SVM (Séparateurs à Vaste Marge). Cette méthode est comparée aux algorithmes de classification plus conventionnels comme les K-moyennes ou les K-plus proches voisins, qui sont régulièrement utilisés en traitement d'image. Nous proposons également une méthode de classification non supervisée combinant les K-moyennes et les SVM. Les résultats fournis par les différentes techniques sont comparés à des vérités de terrain sur diverses zones brûlées.

Mots-clés : feux de forêt, zones brûlées, télédétection, images satellitaires à haute résolution, Séparateurs à Vaste Marge (SVM), K-moyennes, K-plus proches voisins, classification.

Contents

1 Introduction

1.1 Forest Fires

Fires play an important role in forest ecosystems. Each year, several hectares are burnt in the Mediterranean countries, specially during the dry season. For instance, Table (1) shows the average number of forest fires per year and the average burnt area surface in hectares per year for some countries, according to the European Commission (`http://forest.jrc.it`).

	Average	Portugal	Spain	Italy	Greece	France
Burnt area in ha per year	2000-2005	213 972	141 788	71 900	31 581	29 870
	1980-2005	117 335	188 917	115 583	44 413	30 682
Number of fires per year	2000-2005	28 568	21 570	7 739	1 834	4 278
	1980-2005	17 757	15 618	10 603	1 581	5 005

Table 1: Statistics on forest fires.

On one hand, fires are beneficial for forest regeneration. Indeed, they provide a natural way of sterilizing the soil by getting rid of bacteria, dead leaves, branches and shrubs. Moreover, they allow seed dispersal and growths of new and robust plants which replace the aged forest.

On the other hand, wildfires are a major environmental issue in forest ecosystems. They threaten biodiversity by changing the biomass stocks and affecting the forest fauna. Fires can also lead to environmental degradation through impacts on hydrologic cycles, water quality, soil fertility and erosion. Smoke produced by fires can significantly reduce photosynthetic activity and can endanger health of humans and animals.

1.2 Burnt area mapping

Once the fire is extinguished, detailed and accurate information concerning the burnt area location and perimeter is important to assess economic and ecologic losses. This helps to identify the rehabilitation and restoration needs in terms of time and human, material and financial resources. This also determines the risks of flooding, surface erosion and other related processes, such as natural regeneration of vegetation for instance.

1.3 How to assess the burnt areas ?

Several studies have shown the great potential of remote sensing techniques for burnt area mapping [1]. Remote sensing is a valuable tool since it provides spatial information on the entire forest [2], even for remote areas. Moreover, the spectral information given by remote sensing techniques can belong to the visible part of the spectrum (Green and Red channels)

as well as the non-visible part (Near Infra Red and Mid Infra Red channels), which is very relevant for vegetation investigation.

Most of the studies concerning burnt area mapping use multi-temporal series of coarse resolution satellite images taken before and after the fire [3]. They are usually based on the computation of the differences between the different channels of the images or on the differences between a combination of different channels (vegetation indices) of the images [4, 5].

The main interest of our study [6, 7] is to assess burnt areas from a single post-fire high-resolution image. Thus, we do not need a database of images of the same area or to purchase a pre-fire image. Moreover, high resolution satellites which are sun-synchronous, can be used since we do not need daily observation of the same area.

In this report, the discrimination between burnt and unburnt areas is obtained thanks to the radiometric information given by the different spectral bands provided by the sensors of the SPOT 5 satellite. The classification is based on Support Vector Machines (SVM) [8]. SVM are a supervised learning technique well adapted to deal with data of high dimensions such as satellite images. Moreover, they have shown to outperform classical classification algorithms such as the maximum likelihood or multilayer neural networks. For instance, they provide good results in different contexts such as handwritten digit recognition, pattern recognition or biomedical imagery and have been recently introduced in remote sensing image classification [9].

A drawback of SVM is that it requires a learning phase. Indeed, we have to feed it with already labeled objects to learn how to classify new objects. The objects can not be randomly chosen and their choice is crucial for the classification accuracy. This training phase is usually done manually and takes a lot of time. In this report, we propose a combined method incorporating K-means and SVM to help the user to find automatically good training samples.

The performances of SVM and the combined SVM are compared with traditional classification methods such as K-means and K-nearest neighbours algorithms. The maps obtained by each classification are also compared to ground truths derived from manual interpretation of remote sensing images and from ground estimates using GPS (all of them made by experts).

In the following sections, we first give of brief overview of SVM (see Section 2). Experimental results are given in Section 3. Finally, Section 4 concludes this report.

2 Support Vector Machines

Hereafter, we give an introduction about SVM. Interested readers can consult [10] for more details.

The goal of SVM is to use a set of observations to find a decision surface. This surface separates new objects into two classes. Here, we restrict ourselves to the classification into 2 classes (SVM is fundamentally developed for binary classification but can be extended to

multi-class problems). Let R^m be the input space. Each element $\overrightarrow{x} \in R^m$ has to be labeled by a value, denoted by y which can be for instance -1 or $+1$.

Let us assume to be given a set of observations called a training set, $\{(\overrightarrow{x_i}, y_i)\}_{i \in <1,N>}$. where $\overrightarrow{x_i} \in R^m$ and $y_i = \pm 1$. The goal of a learning algorithm is to use this training set to predict the label of all the new objects of R^m. The problem is thus to separate the two classes by finding a good decision function f, which assigns the label $+1$ (respectively -1) to the elements $\overrightarrow{x}$ such that $f(\overrightarrow{x}) \geq 0$ (respectively $f(\overrightarrow{x}) < 0$). The surface defined by the equation $f(\overrightarrow{x}) = 0$ divides the space into two half-spaces, the positive one and the negative one.

An obvious criterion to choose f is to minimize the error rate on the training set, i.e. the number of misclassification on the training set. This principle, called Empirical Risk Minimization (ERM) [11], has some drawbacks: the global error rate can be much higher than the training error rate (it does not usually provide a good generalization ability, which is called overfitting) and it does not define a unique solution, even when it is possible to perfectly separate the training set. The phenomenon of *overfitting* happens when the function which separates the training set has a very complicated shape and is too much related to this training set.

On the contrary of other learning algorithms, SVM are based on the Structural Risk Minimization (SRM) principle [8]. This technique aims at generalizing well rather than correctly classifying the training set and prevent from overfitting. The hyperplane is chosen so as to maximize the *margin*, or twice the distance of the training set from the hyperplane. This hyperplane is called *Optimal Separating Hyperplane* (OSH).

2.1 Linearly separable case

A training set is said to be linearly separable if there exists at least a linear classifier f which correctly classifies all of the elements of the training set. Obviously, this case is very restrictive, but it will allow us to present the main features of SVM. The non-linearly separable case will be studied in the next section.

In the linearly separable case, the OSH (see Figure 1) is defined by the equation $\overrightarrow{w}.\overrightarrow{x} + b = 0$ where $\overrightarrow{w}$ is a vector normal to the hyperplane and b is the bias. Thus, to correctly classify the training set, we must have:

$$\begin{cases} \overrightarrow{w}.\overrightarrow{x_i} + b \geq \epsilon > 0 & \text{if } y_i = +1 \\ \overrightarrow{w}.\overrightarrow{x_i} + b \leq \epsilon < 0 & \text{if } y_i = -1 \end{cases} \quad \text{which is equivalent to} \quad \begin{cases} \overrightarrow{w}.\overrightarrow{x_i} + b \geq 1 & \text{if } y_i = +1 \\ \overrightarrow{w}.\overrightarrow{x_i} + b \leq 1 & \text{if } y_i = -1 \end{cases}$$

These constraints imply that the margin to maximize, defined by the distance between the equations $\overrightarrow{w}.\overrightarrow{x} + b = -1$ and $\overrightarrow{w}.\overrightarrow{x} + b = +1$, is equal to $\frac{2}{||w||}$. The training samples must be on the good side of the tube defined by the equations $\overrightarrow{w}.\overrightarrow{x} + b = -1$ and $\overrightarrow{w}.\overrightarrow{x} + b = +1$.

As maximizing the margin is equivalent to minimizing the inverse of the square margin, the optimal parameters w and b are given by the solution of the convex Quadratic

Programming Problem (QPP) with linear inequality constraints:

$$\min_{(w,b)} \frac{\|\overrightarrow{w}\|^2}{2} \quad \text{subject to: } y_i(\overrightarrow{w}.\overrightarrow{x_i} + b) \geq 1 \ , \ \forall i \in \langle 1, N \rangle \tag{1}$$

By introducing positive Lagrange multipliers $\{\lambda_i\}_{i \in <1,N>}$ associated with the constraints, it leads to the following dual problem:

$$L(\overrightarrow{w}, b, \overrightarrow{\lambda}) = \frac{||\overrightarrow{w}||^2}{2} - \sum_{i=1}^{N} \lambda_i.[y_i(\overrightarrow{w}.\overrightarrow{x_i} + b) - 1] \tag{2}$$

The unique saddle point of the Lagrangian L, which is a minimum with respect to (w, b) and a maximum with respect to λ, determines the solution of the QPP. Therefore, let us first minimize the Lagrangian with respect to (w, b):

$$\frac{\partial L}{\partial \overrightarrow{w}} = \overrightarrow{w} - \sum_{i=1}^{N} \lambda_i y_i \overrightarrow{x_i} = 0 \tag{3}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N} \lambda_i y_i = 0 \tag{4}$$

By substituting these equations into Eq. (2), the dual problem is reduced to the following QPP with bound constraints and a linear equality constraint:

$$\max_{\lambda} W(\lambda) = -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j \overrightarrow{x_i}.\overrightarrow{x_j} + \sum_{i=1}^{N} \lambda_i$$

$$\text{subject to : } \sum_{i=1}^{N} \lambda_i y_i = 0 \ , \ 0 \leq \lambda_i \ , \ \forall i \in \langle 1, N \rangle$$

If we call λ^* the solution of this problem, we determine the solution parameter $\overrightarrow{w}^*$ from (3):

$$\overrightarrow{w}^* = \sum_{i=1}^{N} \lambda_i^* y_i \overrightarrow{x_i} \tag{5}$$

thus b^* can be obtained from the Karush-Kuhn-Tucker (KKT) conditions [8] which derived from the constraints in (1):

$$\lambda_i^*.[y_i(\overrightarrow{w}^*.\overrightarrow{x_i} + b^*) - 1] = 0$$

If λ_i^* is non zero, the corresponding $\overrightarrow{x_i}$ lies exactly on the boundary (at a distance equal to half the margin from the OSH). Otherwise, λ_i^* is null and the corresponding $\overrightarrow{x_i}$ lies on the good side of the boundary.

Note that $\overrightarrow{w}^*$ is a linear combination of the vectors $\overrightarrow{x_i}$ for which the Lagrange multiplier λ_i^* is not equal to 0. These vectors are called *Support Vectors* (SV) and are the closest elements to the OSH. The optimal parameters $(\overrightarrow{w}^*, b^*, \overrightarrow{\lambda}^*)$ do not change if non-support vectors are removed from the training set.

The decision function f is given by:

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

The classification of a new object $\overrightarrow{x}$ requires the computation of the dot product between $\overrightarrow{x}$ and the SV. Moreover, the decision function f is defined completely by the SV and does not depend on the non-SV.

2.2 Non-linearly separable case

In the previous subsection, we assume that the training set was linearly separable and we show how to find the classifier with the best margin. In this subsection, we generalize the linear case to handle non-linearly separable training sets. Indeed, the linear hypothesis is too restrictive for real-life applications and there can be noise, measurement errors or outliers in the training set.

2.2.1 Soft margin case

The previous analysis is extended by introducing *slacks variables* $\{\xi_i\}_{i\in<1,N>}$ in order to allow some classification mistakes in the training set (see Figure 2).

These variables quantify the mistakes made by the training examples. Indeed, The slack variable ξ_i corresponding to the training vector x_i is zero if the vector is on the good side of the tube defined by $\{\overrightarrow{w}, b\}$, otherwise it is equal to the distance between the tube boundary $\overrightarrow{w}.\overrightarrow{x} + b = \pm 1$ and the point x_i.

$$\xi_i(\overrightarrow{w}, b) = \begin{cases} 0 & \text{if } y_i.(\overrightarrow{w}.\overrightarrow{x_i} + b) \geq 1 \\ 1 - y_i.(\overrightarrow{w}.\overrightarrow{x_i} + b) & \text{if } y_i.(\overrightarrow{w}.\overrightarrow{x_i} + b) \leq 1 \end{cases}$$

The soft margin SVM aim at maximizing the margin while minimizing the errors, which implies the new QPP:

$$\min_{(w,b,\xi)} \frac{\|\overrightarrow{w}\|^2}{2} + C \sum_{i=1}^{N} \xi_i \tag{6}$$

$$\text{subject to : } y_i(\overrightarrow{w}.\overrightarrow{x_i} + b) \geq 1 - \xi_i \ , \ \xi_i \geq 0, \ \forall i \in \langle 1, N \rangle \tag{7}$$

where C is called the *regularization parameter*. A larger C assigns a higher penalty on the training error, a smaller C tends to increase the margin and the training errors. This

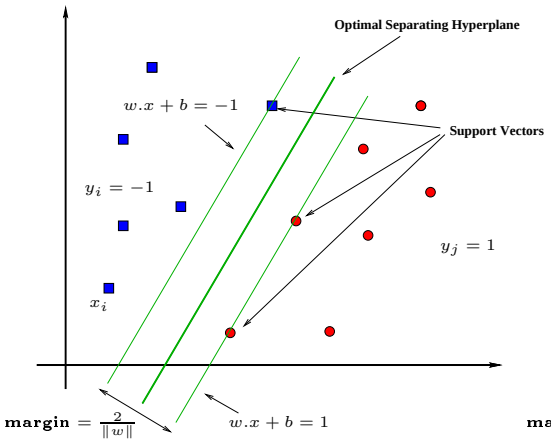


Figure 1: Linear case.

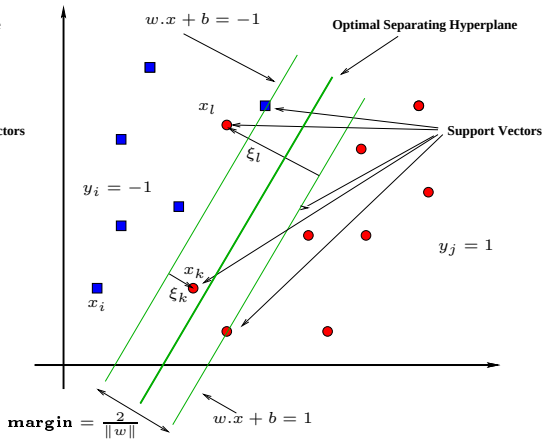


Figure 2: Soft margin case.

parameter defines a trade-off between a larger margin on one hand and a small number of errors on the other hand.

As for the previous primal problem (see Eq. (1)), we introduce Lagrange multipliers $\{\lambda_i\}_{i\in<1,N>}$ and $\{\mu_i\}_{i\in<1,N>}$ associated respectively with the constraints $y_i(\overrightarrow{w}.\overrightarrow{x_i} + b) \geq 1 - \xi_i$ and $\xi_i \geq 0$:

$$L(\overrightarrow{w}, b, \overrightarrow{\xi}, \overrightarrow{\lambda}, \overrightarrow{\mu}) = \frac{||\overrightarrow{w}||^2}{2} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\lambda_i.[y_i(\overrightarrow{w}.\overrightarrow{x_i} + b) - 1 + \xi_i] - \sum_{i=1}^{N}\mu_i.\xi_i \tag{8}$$

To minimize the Lagrangian with respect to $\{\overrightarrow{w}, b, \overrightarrow{\xi}\}$, let us compute the partial derivatives:

$$\frac{\partial L}{\partial \overrightarrow{w}} = \overrightarrow{w} - \sum_{i=1}^{N}\lambda_i y_i \overrightarrow{x_i} = 0 \tag{9}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{N}\lambda_i y_i = 0 \tag{10}$$

$$\frac{\partial L}{\partial \xi_i} = C - \lambda_i - \mu_i = 0 \tag{11}$$

By substituting these equations into Eq. (8), we obtain:

$$\max_{\lambda} W(\lambda) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \overrightarrow{x_i}.\overrightarrow{x_j} + \sum_{i=1}^{N}\lambda_i$$

$$\text{subject to : } \sum_{i=1}^{N}\lambda_i y_i = 0 \ , \ 0 \leq \lambda_i \leq C \ , \ \forall i \in \langle 1, N\rangle$$

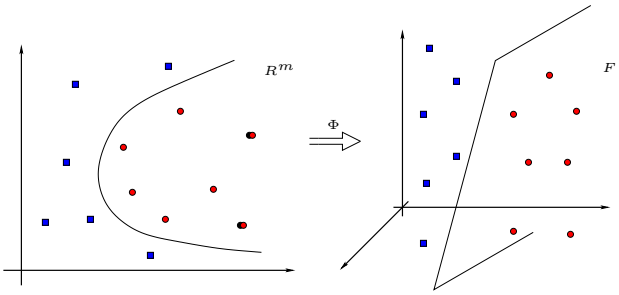


Figure 3: Role of the mapping Φ.

This optimization problem is the same as before, except that now, λ has to be upper bounded. Thus, after having solved this problem, we recover the solution $\overrightarrow{w}^*$ from the same equation (5) as before .

Note that the KKT conditions which allow to compute b are:

$$\lambda_i.[y_i(\overrightarrow{w}.\overrightarrow{x_i}+b)-1+\xi_i]=0$$

$$\mu_i.\xi_i=0$$

which means that:

- if $\lambda_i = 0$, then $\mu_i = C$, $\xi_i = 0$ and $[y_i(\overrightarrow{w}.\overrightarrow{x_i}+b)-1] > 0$; the vector x_i is on the good side of the tube,
- if $0 < \lambda_i < C$, then $0 < \mu_i < C$, $\xi_i = 0$ and $[y_i(\overrightarrow{w}.\overrightarrow{x_i}+b)-1] = 0$; the vector x_i lies on the boundaries of the tube, and is a SV,
- if $\lambda_i = C$, then $\mu_i = C$, $\xi_i > 0$ and $[y_i(\overrightarrow{w}.\overrightarrow{x_i}+b)-1+\xi_i] = 0$; x_i is not on the good side of the tube (either well classified inside the tube and in the right side of the OSH, or badly classified on the other side of the OSH), but it is still a SV since it contributes to the definition of $\overrightarrow{w}$ and also to the position of the OSH.

2.2.2 Non-linear SVM

In most cases, the training set needs to be separated by a more complicated shape (a circle for example). To handle a non-linear OSH, the linear SVM are generalized by introducing a function Φ that maps the data of the input space R^m into a higher dimensional feature space F where the data are linearly separable (see Figure 3):

$$\begin{array}{rcl}\Phi : R^m & \longrightarrow & F \\ \overrightarrow{x} & \longmapsto & \Phi(\overrightarrow{x})\end{array}$$

If we find a mapping Φ such that the new training set $\{(\Phi(\overrightarrow{x_i}), y_i)\}_{i\in<1,N>}$ is linearly separable, by analogy with the previous analysis, the OSH parameters are given by the solution of the following QPP:

$$\max_{\lambda} W(\lambda) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j y_i y_j \Phi(\overrightarrow{x_i}).\Phi(\overrightarrow{x_j}) + \sum_{i=1}^{N}\lambda_i \tag{12}$$

$$\text{subject to : } \sum_{i=1}^{N}\lambda_i y_i = 0 \ , \ 0 \leq \lambda_i \leq C \ , \ \forall i \in \langle 1, N\rangle \tag{13}$$

which gives

$$\overrightarrow{w}^* = \sum_{i=1}^{N}\lambda_i^* y_i \Phi(\overrightarrow{x_i}) \tag{14}$$

$$f(\overrightarrow{x}) = \sum_{i=1}^{N}\lambda_i^* y_i \Phi(\overrightarrow{x_i}).\Phi(\overrightarrow{x}) + b^* \tag{15}$$

These formulas are obtained by substituting the input vector $\overrightarrow{x}$ by its value in the feature space $\Phi(\overrightarrow{x})$.

Note that, in Eq. (12) and (15), Φ contributes only to the dot product between feature examples. All we need to know is the value of the dot product for any couple of points of the input space, which can be called *Kernel*:

$$K(\overrightarrow{x}, \overrightarrow{x}') = \Phi(\overrightarrow{x}).\Phi(\overrightarrow{x}')$$

Consequently, if we find this expression, then the explicit knowledge of Φ is not necessary. This is the *kernel trick* [12].

A useful criterion to know if a function can be a valid kernel is given by Mercer theorem: a symmetric function K is a kernel if and only if $K(x_1, x_2)$ is a positive definite matrix. This theorem allows us to compute directly the dot product $K(\overrightarrow{x_1}, \overrightarrow{x_2})$ without computing the images $\Phi(\overrightarrow{x_1})$ and $\Phi(\overrightarrow{x_2})$. Thus the expensive computation of Φ is avoided, as well as its explicit definition.

Note that the kernel corresponds to a dot product (symmetric definite positive function) in the feature space. The optimization is still a convex QPP with linear equality constraints since the kernel is positive definite.

2.2.3 Some kernels

Kernels characterize the similarity degree or the correlation between two samples. They must have high values for closer samples and low values for further samples. The most classical kernels (which are widely used in remote sensing) are:

- the linear kernel: $K(\overrightarrow{x}, \overrightarrow{x}') = \overrightarrow{x}.\overrightarrow{x}'$, which is used in the linear case.
- the polynomial kernel: $K(\overrightarrow{x}, \overrightarrow{x}') = (\overrightarrow{x}.\overrightarrow{x}' + 1)^q$. The larger the degree q, the more complicated the OSH.
- the RBF (Radial Basis Function) kernel: $K(\overrightarrow{x}, \overrightarrow{x}') = \exp(-\frac{\|\overrightarrow{x}-\overrightarrow{x}'\|}{\sigma})$.
- the Gaussian kernel: $K(\overrightarrow{x}, \overrightarrow{x}') = \exp(-\frac{\|\overrightarrow{x}-\overrightarrow{x}'\|^2}{2\sigma^2})$.

Almost every shape can be obtained with the last two kernels. The smaller the parameter σ is, the more complicated the OSH is. These kernels correspond to a mapping in a feature space of infinite dimension.

Other kernels have demonstrated good classification results:

- the KMOD (Kernel with MODerate decrease) kernel [13]: $K(\overrightarrow{x}, \overrightarrow{x}') = a[\exp(\frac{\gamma}{\|\overrightarrow{x}-\overrightarrow{x}'\|^2+\alpha^2})-1]$ with $a = \frac{1}{\exp(\frac{\gamma}{\alpha^2})-1}$. This kernel exhibits a quick decrease near zero and a moderate decrease near infinity. It tends to discriminate much better close samples and to preserve some information for distant samples.

- the SAM (Spectral Angle Mapper) kernel [14]: $K(\overrightarrow{x}, \overrightarrow{x}') = \exp(-\gamma \arccos(\frac{\overrightarrow{x}.\overrightarrow{x}'}{\|\overrightarrow{x}\|.\|\overrightarrow{x}'\|}))$. This kernel is scale invariant and does not depend on a quadratic distance. It seems to be relevant for image classification since it is more robust to differences of illuminations or shadows between two vectors.

Note that new kernels can be obtained from simple operations on kernels. Indeed, a linear combination of kernels is a valid kernel (it fulfills Mercer condition). Thus we can benefit from advantages of two kernels by combining them:

$$K(\overrightarrow{x}, \overrightarrow{x}') = \mu.K_{Gaussian}(\overrightarrow{x}, \overrightarrow{x}') + (1-\mu).K_{SAM}(\overrightarrow{x}, \overrightarrow{x}')$$

2.2.4 Implementation

There are several standard methods to solve the convex Quadratic Programming Problem with linear and box constraints (see Eq. (12) and (13)) such as the conjugate gradient method. Nevertheless, these methods are not very effective when the number of training samples is important. Specialized algorithms have been developed for quickly solved the QPP.

The chunking algorithm proposed by Vapnik [8] uses the fact that the solution of the QPP is the same if we remove non-SV. This algorithm breaks the large QPP into a series of smaller QP sub-problems and aims at identifying the non-zero Lagrange multipliers and discarding the zero Lagrange multipliers.

A common and simple method to solve QPP is the Platt Sequential Minimal Optimization (SMO) algorithm [15] which breaks the QPP into a 2-dimensional sub-problem which can be solved analytically and then avoid the time-consuming numerical quadratic optimiza-

tion.

Nevertheless, there are no criteria concerning the choice of the parameters involved in the QPP (parameter C of the soft margin case and kernel parameter σ for the Gaussian kernel for instance). A way to choose these parameters is to take several values and test their efficiency in order to select the most suited one, i.e. by *trial and error*. Another usual way is the *cross-validation* technique [12]. For each parameter, we first divide the training set into several subsets and test sequentially the obtained classification on one of them after having trained SVM on the other subsets.

3 Experiments

3.1 Data

The data used in the experiments are SPOT 5 satellite images representing Southern France areas. Either the images are at 10-meter-resolution and consist in the Green, Red, Near Infra-Red (NIR) and Mid Infra-Red (MIR) bands or they are at 2.5-meter-resolution but consist only in the Green, Red and the NIR bands. These latter images are the result of the fusion from a 2.5 meter-resolution panchromatic SPOT 5 image and the corresponding 10-meter resolution multi-spectral SPOT 5 image.

The table below gives some information concerning the data: the area, the burnt surface in hectares, the year, the approximate time lag between the extinction of the fire and the acquisition of the SPOT 5 image used in the experiments, the satellite image resolution and the type of ground truth, i.e. if they derived from Manual Field Checking (MFC) or Aerial Interpretation (AI, in this case the image resolution is given), or if we have both.

n^o	Area	ha	year	time lag	resolution	ground truth
1	Massif des Maures	19.000	2003	1 day	2.5 m	AI - 2.5m
2	Luceram	2.000	2003	2 months	10 m	MFC
3	Northern Corsica	20.000	2003	1 day	10 m	MFC
4	Western Corsica	2.000	2005	1 month	10 m	MFC and AI - 30m
5	Eastern Corsica	2.000	2005	1 month	10 m	MFC and AI - 30m
6	Alpes-Maritimes	500	2007	1 day	10 m	MFC and AI - 30m

Table 2: Statistics on Forest Fires.

The ground truth of image n^o 1 was provided by the French Space Agency (CNES) and derived from the corresponding SPOT 5 satellite image. The ground truth of the image n^o 2 was provided by the French Forestry Office (ONF). For the other images, the SDIS2B (Service Départemental d'Incendie et de Secours de Haute-Corse) provided us with the ground truths of the largest burnt areas and Infoterra France - ESA provided us with the ground

truth derived from Landsat satellite images (30 meters of resolution).

For instance, Figure 4 shows the false colour composite of a SPOT 5 multispectral image (NIR, Red and Green), the four bands separately and the corresponding ground truth. Vegetation appears in red (higher reflectance in the NIR band), sea in blue (higher Green band reflectance), clouds in white (same reflectance in the three bands). Bare soils, roads and urban areas appear in various shades of blue, yellow or grey, depending on their composition. Burnt areas appears in dark grey and can be seen more easily in the NIR band than in the other three bands.

Several studies based on the change detection between pre-fire and post-fire images consider the differences between different channels of the satellite image or between relevant combinations of different channels called *vegetation indices* [16].

The most popular vegetation index is the NDVI (Normalized Difference Vegetation Index) which measures the density of vegetation for a pixel: $NDVI = \frac{NIR-Red}{NIR+Red}$. Another vegetation index which is used for burnt area mapping is the NDII (Normalized Difference Infra-red Index), also called NBR (Normalized Burn Ratio) which is obtained by replacing the red reflectance value with the Mid infra-red reflectance value: $NDII = \frac{NIR-MIR}{NIR+MIR}$.

When using several images, vegetation indices are very useful since they tend to emphasize the spectral change of vegetation and to be less sensitive to the illumination difference of the two images. Nevertheless, when using only the post-fire image, there is no use thresholding spectral bands or vegetation indices to detect the burnt areas. Figure 5 shows the values of the NIR, NDVI and NDII for the satellite image represented in Fig. 4 and the corresponding classifications obtained after thresholding.

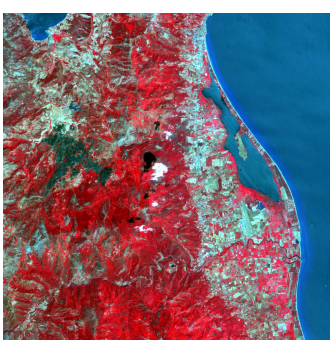
In the following, if the data are not specified, the image of the "Massif des Maures" (2003) will be used to avoid the registration problems when comparing the ground truth with the extracted classifications.

3.2 SVM Settings

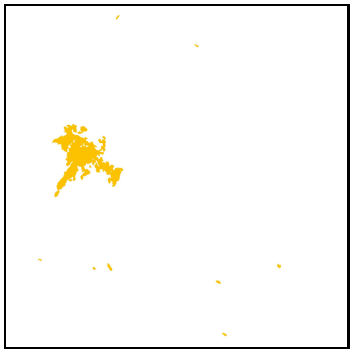
3.2.1 Training set

The training set needs to provide a complete and representative description of the two classes. Several iterations are needed to manually select a good one [17]. At each iteration, SVM are launched on a part of the image and the most characteristic misclassified pixels are added to the training set. As the classification result depends on the SVM parameters, some pixels which have just been added can be misclassified again at the next iteration and thus do not need to be added again (they could be well-classified with other parameters).

Figure 6 shows the overall accuracy and the computing time of the SVM and K-nearest neighbours (K-NN) algorithms with respect to the ground truth according to the training



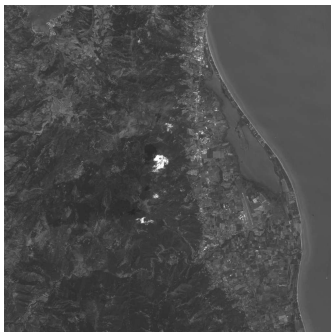
(a) False color image



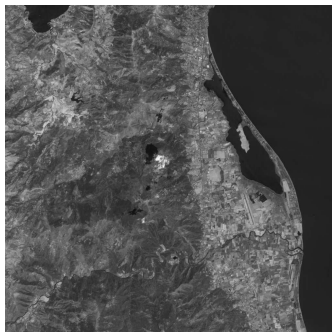
(b) Ground Truth

(c) Near Infra Red band

(d) Red band



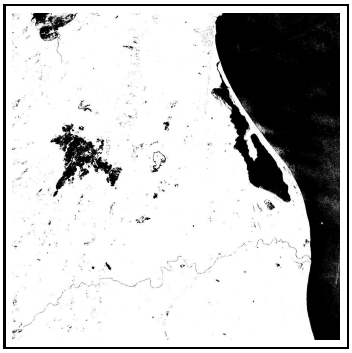
(e) Green band



(f) Mid Infra Red band

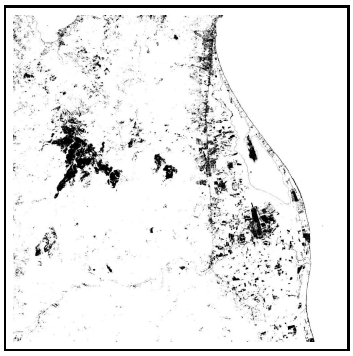
Figure 4: Eastern Corsica - SPOT 5 image © CNES 2005, Distribution SPOT Image and Ground Truth - projet RISK-EOS © Infoterra - ESA.

(a) NIR band

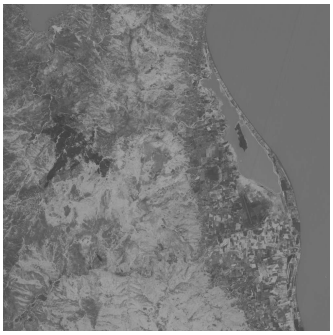


(b) NIR band + threshold

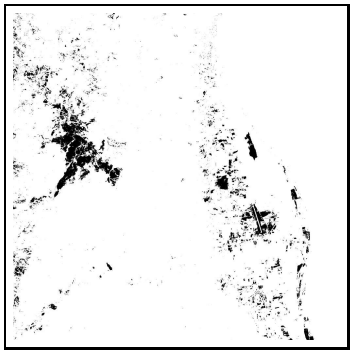
(c) NDVI



(d) NDVI + threshold



(e) NDII



(f) NDII + threshold

Figure 5: Eastern Corsica - SPOT 5 image © CNES 2005, Distribution Spot Image.

set percentage. We applied these algorithms on a 3-band image (10.000*5.000 pixels). The total training set size was equal to 550 (220 for the burnt class and 330 for the unburnt one). The computer used was a 2 GHz processor with 2 Go RAM.

Note that the computing time is proportional to the number of elements of the training set for the K-NN and to the number of Support Vectors for the SVM (see Eq. (15)). For the SVM, the training set size does not play an important role but it is the relevance of the training samples which gives a good classification.

In this experiment, the hyperparameters of SVM are the same for all the training sets. A better result could be obtained for the first training set but will need more SV. Except for the smallest training set, SVM outperform K-NN concerning the overall accuracy but are slower.

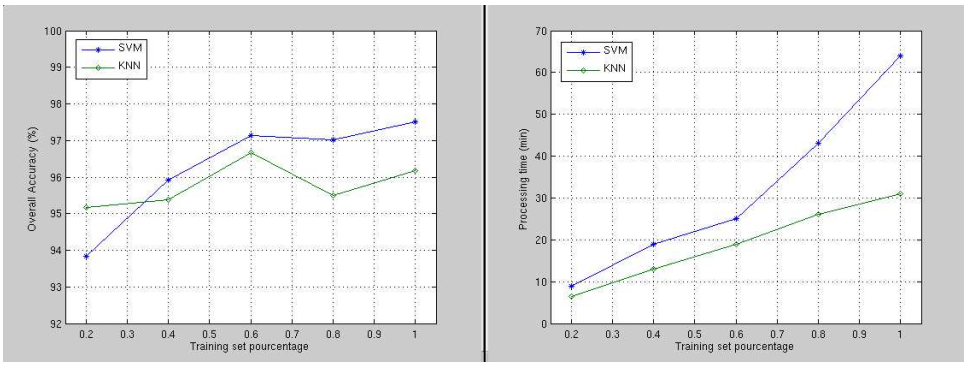


Figure 6: Overall accuracy and computing time.

3.2.2 Kernel choice

A keypoint of SVM is the choice of the kernels and their parameters. No single kernel works correctly for all the problems. Several kernels have been considered in this report. For each kernel, different parameters have been tested to determine empirically the best ones. Table (3) provides the overall accuracy (OA), the False Positive rate (FP) and False Negative rate (FN) between the extracted burnt areas and the ground truth.

According to this table, concerning the kernel choice, as the different kernels give approximately the same results, we choose to use the Gaussian one in the next sections. This kernel is the most popular since it is very simple, depends on a single parameter and does not need a high computing time. Another advantage lies in the fact that it maps the data into a infinite dimension space, which makes possible to separate any training set linearly on the contrary to polynomial kernels.

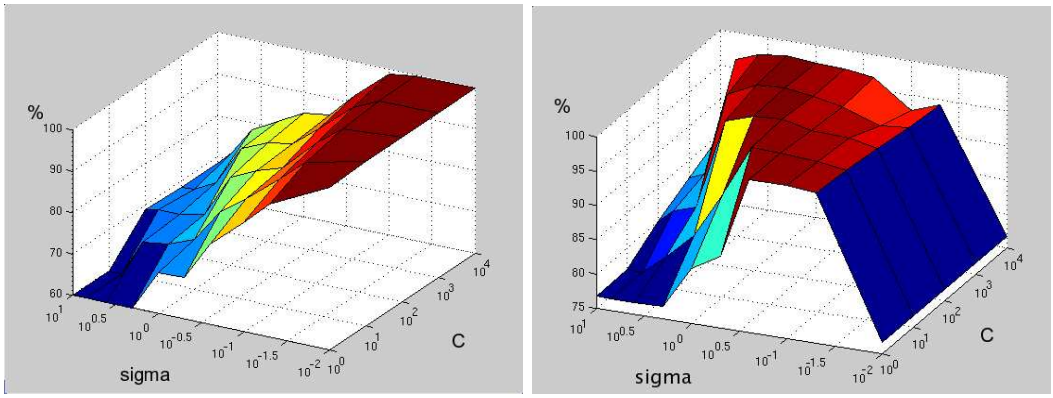
Ground Truth (2550 ha)	Burnt area (ha)	OA (%)	FP (%)	FN (%)
SVM pol.	2238	98.50	3.03	1.37
SVM G.	2279	98.73	2.41	1.18
SVM SAM+G.	2084	98.14	2.13	1.84
SVM KMOD	2045	98.05	1.90	1.96

Table 3: Performances of the different classifiers.

3.2.3 Hyperparameter setting

The kernel parameters and the regularization parameter are called the *hyperparameters* of the SVM classifier.

Experimental results: Figure 7 shows the overall accuracy between the SVM classification and the ground truth according to the different hyperparameters (C for the soft margin case and σ for the Gaussian kernel). According to this figure, several hyperparameter values give a very good agreement between the resulting SVM classification and the ground truth.



(a) Training accuracy (b) Overall accuracy

Figure 7: Hyperparameter setting.

The kernel parameter σ corresponds to the standard deviation of the Gaussian kernel. Let us study more accurately the influence of each parameter on the training and overall accuracies. First, the parameter C is set to the value 100 and let us increase σ:

- If σ is too small, the number of SV is high and the classifier shape is very complicated since it is too much related to the SV. It is the overfitting phenomenon: fitting exactly

to the training data can lead to a model which characterizes very well the training set but will not be representative of the complete input space. The training error rate is almost null but the global error rate is high.

- If σ increases, the classifier shape becomes less complicated, the overall accuracy increases. The training error rate is not null but the global error rate is better.
- If σ is too large, the classifier shape becomes too simple and does not take into account the training data information which results in a low global accuracy. The training error rate is not null and the total error rate is high.

The regularization parameter C defines a trade-off between minimizing the training set error and maximizing the margin, i.e. it controls the balance between the training accuracy and the classifier complexity. Now, the parameter σ is set to the value 1 and let us increase C:

- If C is too small, we allow too many misclassified training samples. The training and overall accuracies are bad since the classifier dismisses training set information.
- If C increases, we penalize more the training error and the training accuracy increases.
- If C is too large, we prevent a high training error, which can lead to the overfitting phenomenon. The overall accuracy can decrease. Nevertheless, when C tends to infinity, the overall accuracy tends to be constant since the maximum of the quadratic programming problem is obtained (see equation (12)).

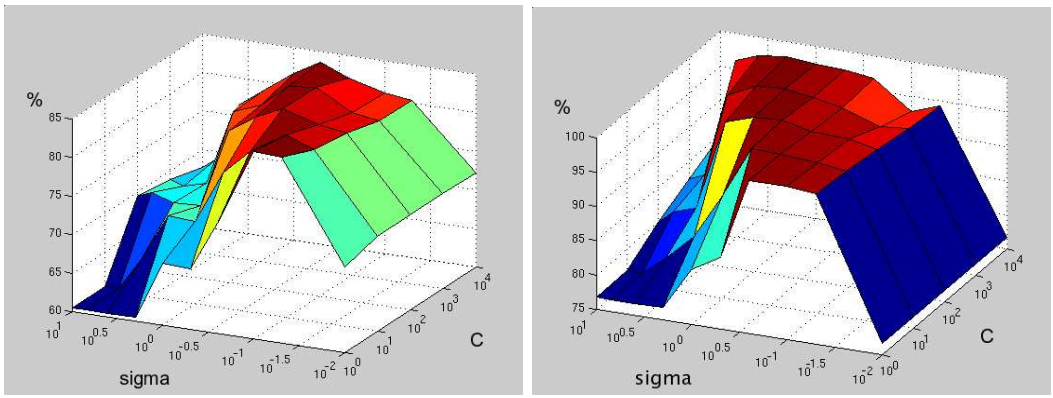
Cross-validation technique: A well-known way for the parameter choice is the *K-fold cross-validation* technique [12]. It consists in dividing the original training set in K subsets, training the SVM on K-1 subsets and testing the resulting classification on the remaining subset. This process is repeated K times, with each of the K subsets. The K results are then averaged to produce the test error. The selected hyperparameters are the ones which yield to the best cross-validation accuracy.

Figures 8 shows the accuracy of the K-fold cross-validation technique and the overall accuracy according to the hyperparameters. The cross-validation technique gives almost the optimal hyperparameters. Anyway, the optimal hyperparameter values may not be the same for both accuracies since minimizing the training error is not equivalent to minimizing the overall error.

If the number of folds K is large: the test set will be small, the test bias will be small but the test variance will be large and the computing time will be very large as well.

If the number of folds is small: the computing time will be reduced as well as the variance of the estimator but the test bias will be large (the number of training samples may not be sufficient comparing to the number of test samples)

In practice, the choice of the number of folds depends on the size of the dataset. For large datasets, even 3-Fold Cross Validation will be quite accurate. For very sparse datasets, we may have to use more folds in order to train on as many examples as possible.



(a) Cross-validation accuracy (b) Overall accuracy

Figure 8: Hyperparameter setting.

3.2.4 Radiometric information

Figure 9 gives the shape of the Optimal Separating Hyperplane (OSH), the position of the training set (left) and the position of all the pixels belonging to the input image (right). In the left figure, the green and blue stars correspond to Support Vectors (respectively for the unburnt and burnt training set) and the green and blue circles correspond to non-SV.

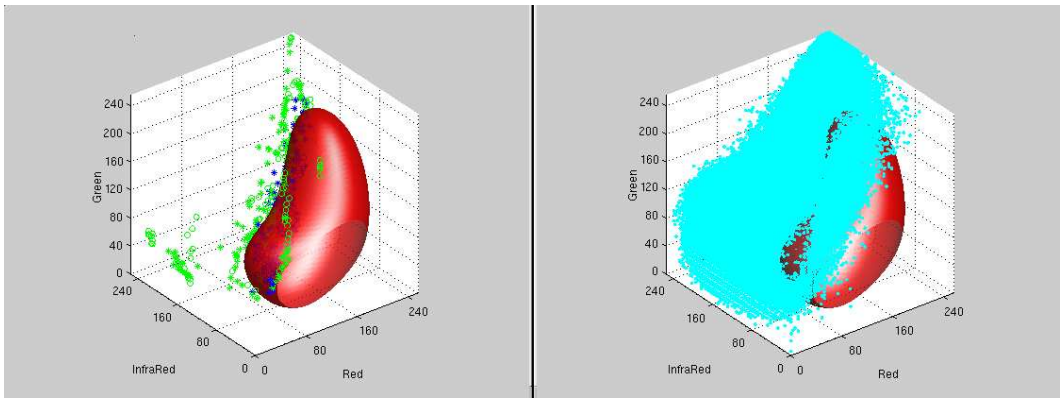


Figure 9: Optimal Separating Hyperplane (OSH).

Figure 10 gives the position of False Positives (left), False Negatives (middle) and True Positives (right) w.r.t. the OSH shape. Pixels considered as burnt in the ground truth belong to a large part of the 3-dimension space and several colours occur in the three figures.

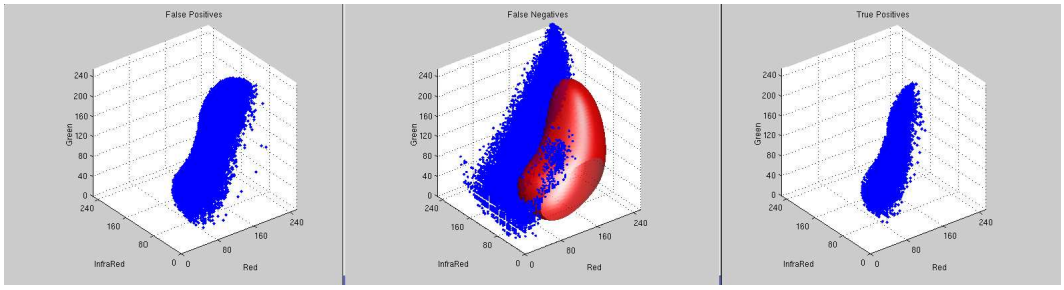


Figure 10: FP, FN and TP.

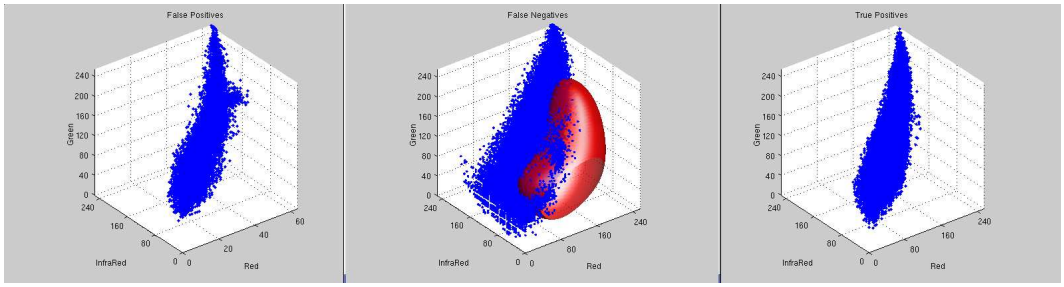


Figure 11: FP, FN and TP after regularization.

3.3 Classification techniques

In this study, we apply the SVM classification technique with a soft margin to satellite images for burnt area mapping. We also proposed an unsupervised method combining K-means and SVM. We compare the obtained results to those given by two traditional classification algorithms which are widely used in remote sensing: the K-means and the K-nearest neighbours. For each classifier, the discrimination between burnt and unburnt areas is first based on the spectral signature.

Regularization: Nevertheless, after classification, the extracted burnt and unburnt areas present some very small discontinuities, i.e., isolated pixels. Therefore, the classification has to be homogenized by reclassifying these pixels.

As burnt and unburnt areas are defined by a set of homogeneous and connected pixels, it is relevant to improve the classification obtained by each classifier by regularizing it. We use either a classical Markovian model (with Potts prior) [18, 19] or mathematical morphology [20]. However, the regularization modifies the shape of pixel sets by smoothing them.

Figure 11 gives the position of False Positives (left), False Negatives (middle) and True Positives (right) after regularization w.r.t. the OSH shape. Note that, after regularization, the TP are not included into the OSH only.

3.3.1 Classical classification techniques

The K-means classifier [21], which is the most frequently used classification technique, is an automatic method which discriminates the input data into K classes. It starts by randomly choosing K centers of gravity, called centroids. Then, it assigns a sample to the class whose centroid is the nearest w.r.t. the Euclidean distance. Afterwards it recomputes the center of gravity of each class. The two last steps are repeated until all the centroids do not change. Finally the K classes are manually merged into two sets to form the burnt and the unburnt areas.

The K-nearest neighbours [21] is a supervised learning method which is commonly used in pattern recognition problems as a reference classification method. It assigns a sample to the most represented class among the K nearest samples of the training set. In our experiments, it is fed with the same training set as the SVM.

3.3.2 An unsupervised algorithm combining K-means and SVM

The SVM algorithm, as well as other supervised learning techniques, requires a set of labeled samples (the training set) to predict the classification of any unlabeled ones. The manually training set selection is crucial and costly since several training samples for each class (burnt areas, forest, fields, sea, lakes, urban areas, roads...) have to be specified to optimize the classification result [21]. Note that, for each class, the spectral characteristics can depend on the slope orientation. Moreover, each image requires its own training set since spectral characteristics of objects (burnt areas, vegetation, urban areas, water, ...) differ from an image to another.

To avoid this key-problem, we propose a process to automatically select the SVM training set. Firstly, the K-means algorithm is used to give a primary label to each pixel. This divides the image in K classes, some corresponding to the burnt areas, the other ones to the unburnt areas. An erosion process is applied to each class in the satellite image in order to eliminate the misclassified pixels and to keep only the pixels which are closer to the center of the connected pixel sets. Finally, these remaining pixels constitute the SVM training set.

3.4 Results

3.4.1 Comparison of the four techniques

Afterwards, to differentiate the two algorithms using SVM, the term "simple SVM" will stand for the SVM classification with a manually chosen training set and the term "K-means + SVM" will denote the unsupervised algorithm combining K-means and SVM introduced in the above section. Both of them use a soft margin and Gaussian kernel.

The figure 12 shows a part of the SPOT 5 image of the "Massif des Maures", the associated ground truth and the four extracted burnt areas.

Table 4 provides the overall accuracy (OA), the True Positive rate (TP - rate of good detection of the pixels classified as burnt in the ground truth), the False Positive rate (FP) and False Negative rate (FN) between the extracted burnt areas and the ground truth.

According to this table, the simple SVM shows a higher discrimination ability than the two other methods. For example, the lakes and the seashore, which have spectral characteristic similar to burnt areas are classified as burnt by the two conventional algorithms. SVM are more sensitive to features with closer spectral characteristics and distinguish easily lake from scar fires.

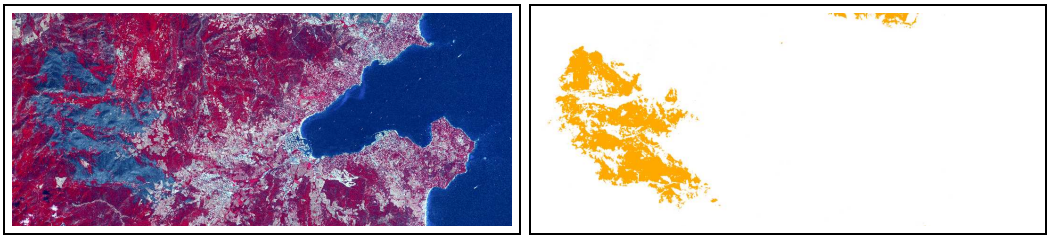
Figure 13 shows a zoom on the differences between the ground truth and the extracted burnt areas.

For the simple SVM classification, the difference map is essentially composed by pixels which lie on the boundaries of the burnt areas. The burnt pixels on the contour have been less damaged by fire than the other ones and have spectral characteristics more similar to healthy vegetation. Moreover, the regularization smooths the contours. For the other classifications, the difference maps are also composed by small pixels sets.

Concerning the supervised algorithms, although the K-NN and the simple SVM use the same training set, the former classification technique shows a higher accuracy than the latter. It has a higher generalization ability. Concerning the unsupervised algorithms, the K-means+SVM gives better result than the K-means only.

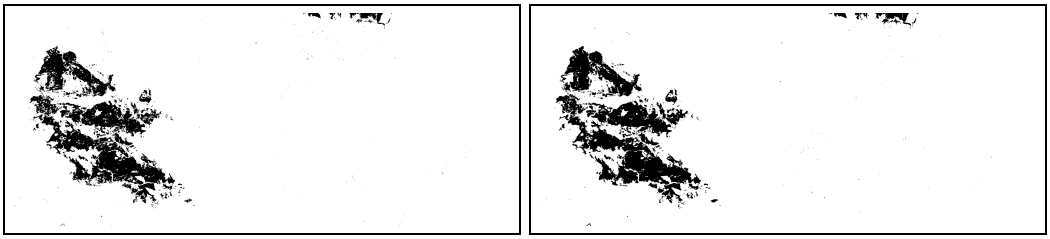
The simple SVM, for which the training set was manually chosen gives better result than the K-means+SVM.

Depending on the hyperparameter choice and on the regularization applied after classification, the SVM can give a better TP rate and a better FN rate. Nevertheless, the OA is lower on the contrary to the FP (see Table 5). When choosing the hyperparameters and the regularization, there is a trade-off between high OA/FP and high TP/FN rates.



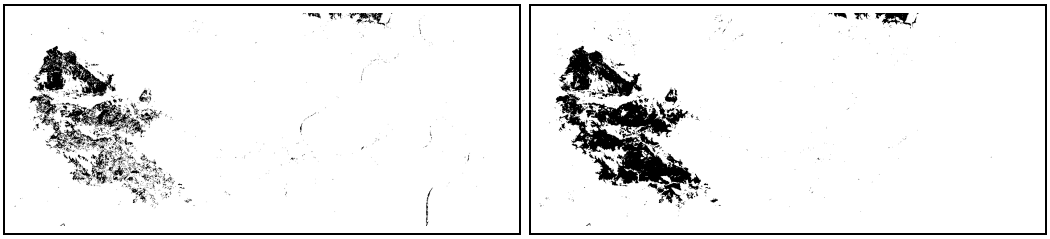
(a) SPOT 5 image © CNES 2003, Distribution SPOT Image

(b) Ground Truth © CNES



(c) K-means

(d) K-means + SVM



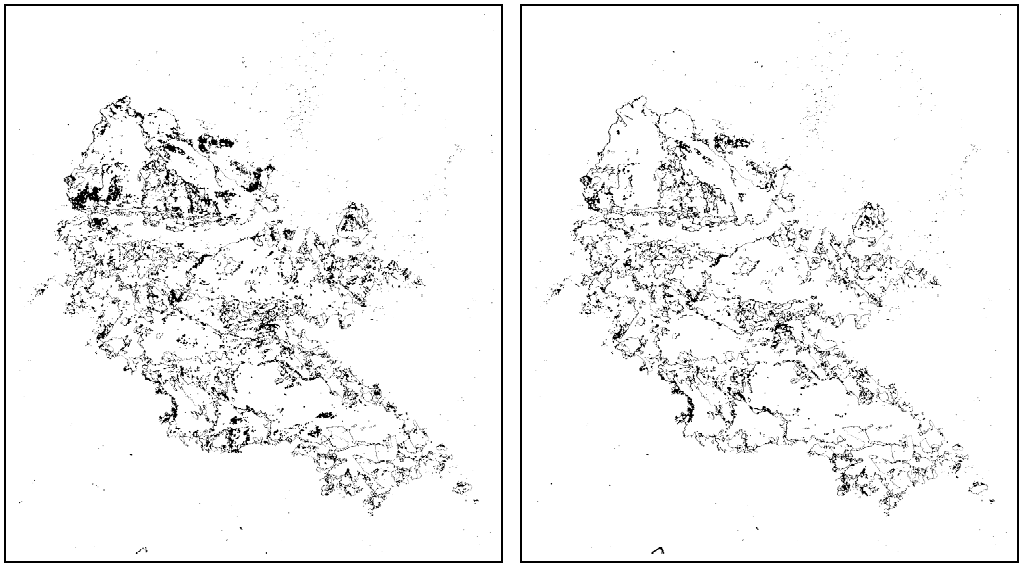
(e) K-NN

(f) simple SVM

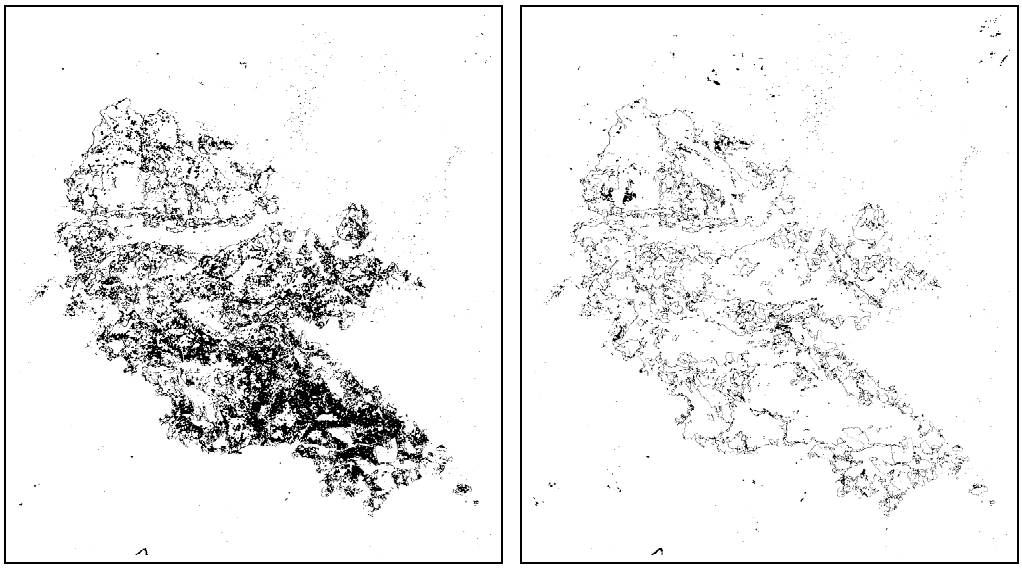
Figure 12: Data and experimental results obtained by different techniques.

Ground Truth (2550 ha)	Burnt area (ha)	OA (%)	TP (%)	FP (%)	FN (%)	time (min)
K-means	1967	97.90	76.26	1.12	2.17	45
K-NN	1366	95.75	51.90	3.09	4.31	30
simple SVM	2343	98.81	89.02	3.14	1.02	60
K-means + SVM	2100	98.39	81.64	0.90	1.69	90

Table 4: Performances of the different classifiers.



(a) K-means (b) K-means + SVM



(c) K-NN (d) simple SVM

Figure 13: Difference maps between the ground truth and the burnt areas.

Ground Truth (2550 ha)	Burnt area (ha)	OA (%)	TP (%)	FP (%)	FN (%)	time (min)
case 1	2343	98.81	89.02	3.14	1.02	60
case 2	2573	98.59	92.23	8.59	0.73	60

Table 5: Performances of SVM with different hyperparameters and different post-processing (case 1: opening then closing with disk of radius 4 as structuring element, case 2: opening then closing with disk of radius 2 as structuring element)

3.4.2 Robustness

Test on other images: In this subsection, we apply the SVM on other images to test their efficiency according to the different burnt vegetation types (moor, garrigues, scrub, broad-leaved tree, coniferous tree) and the area topography (plain, mountain).

The figures concerning the SVM setting (cf. section 3.2) are not displayed for the other images since they have the same shapes.

Concerning the training set choice (when it is manually chosen), its size varies from 300 to 500 samples and it is composed of connected sets of 10 pixels in the image (the total image size is about 6000*6000). These connected sets describe the different classes in the image.

Concerning the hyperparameter choice, the optimal parameters are not the same for all the images but several couples provide good results for one image. These couples are in the range [0.1,1] for σ and vary from 10 to 1000 for C (which is also the range of values given by cross-validation).

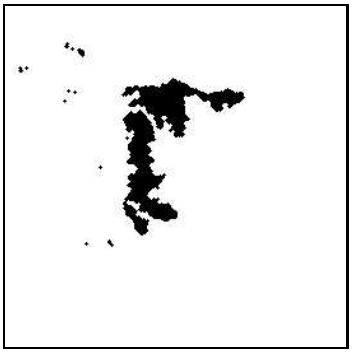
The figures below (see Fig. 14, 15, 16, 17, 18, 19) show some pieces of the different satellite images, the corresponding ground truth(s) and the extracted burnt areas.

Detection of small burnt areas: Up to now, we have concentrated our efforts on the extraction of burnt areas whose surface exceeds 10 hectares since it is the minimal size of the burnt areas ground truth provided by Infoterra France - ESA and SDIS. In this section, we also investigate the assessment of small burnt areas (between 1 and 10 hectares).

The ground truths given by the SDIS2B concern only large burnt areas in Corsica and were obtained manually. The ground truths given by Infoterra France - ESA were derived from the interpretation of Landsat images at 30 meters of resolution, which explains the fact that areas smaller than 10 hectares can not be always detected. However, the Prométhée database (`http://www.promethee.com`) provides some information (date, area coordinates, surface burnt, vegetation type,...) concerning all the fires in France. Thus, it is possible to correlate real fires with extracted burnt areas and determine whether an extracted burnt area corresponds to a real fire.

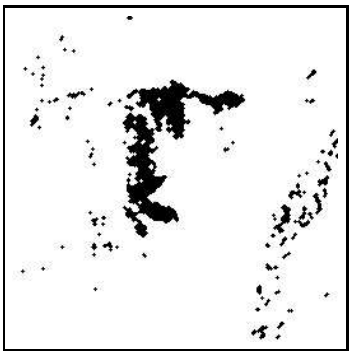
(a) SPOT 5 image © CNES 2003, Distribution SPOT Image

(b) Ground Truth - © ONF-AM



(c) SVM

(d) K-means



(e) K-NN

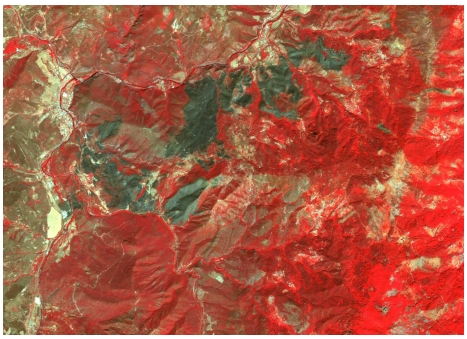
Figure 14: Zoom of Luceram - 2003 (about 50 hectares).

Let us consider Figure 16 which represents a part of Eastern Corsica.

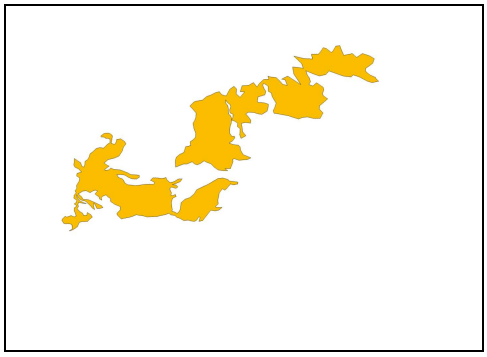
The ground truth provided by Infoterra France - ESA (see Figure 16-b) shows 9 burnt areas. The SVM technique (see figure 16-d) extracts all of the real burnt areas which are in the ground truth, even the smallest ones (cf. burnt area n^o1 - 3 hectares - in Figure 20 which represents a zoom of the previous image). Moreover it also extracts some real burnt areas which are not in the ground truth but which are in the Prométhée database (cf. burnt area n^o2 in Figure 20).

Nevertheless, the extracted and real burnt area boundaries are not the same (due to the regularization, cf section 3.3). Some extracted burnt areas correspond to false alarms (cf. burnt area n^o3 in Figure 20) and all the burnt areas smaller than 10 hectares and belonging to the Prométhée database are not detected.

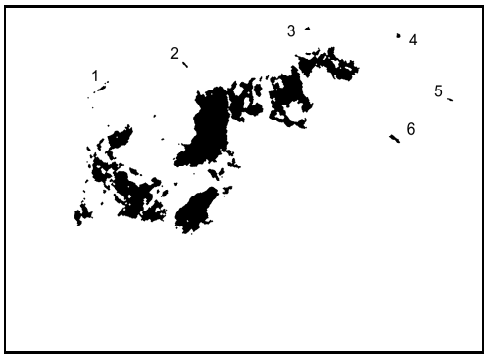
Note that in Figure 15, only the burnt area n^o 1 does not correspond to a real fire. The other ones, whose surface is between 1 to 5 hectares correspond to real fires.



(a) SPOT 5 image © CNES 2003, Distribution SPOT Image



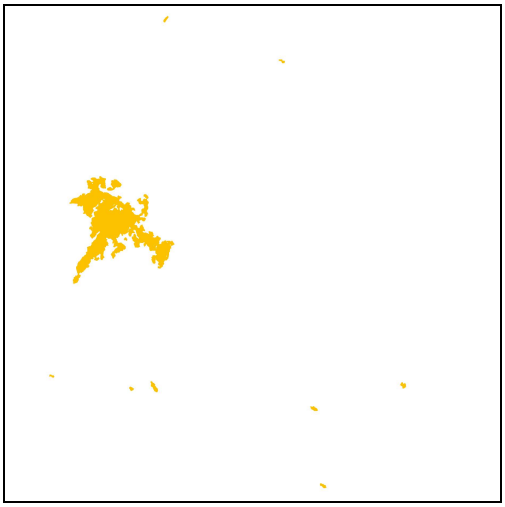
(b) Ground Truth - © SDIS2B



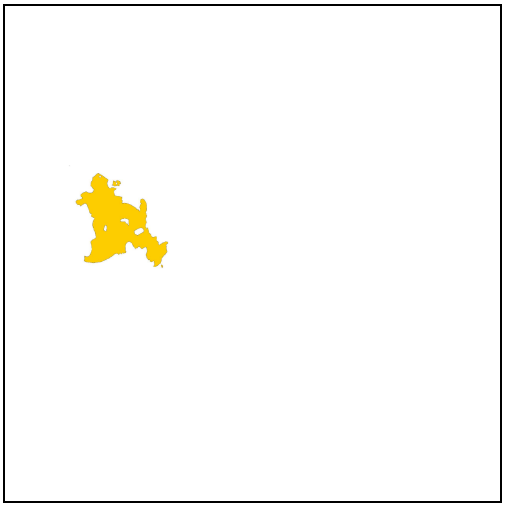
(c) SVM

Figure 15: Zoom of Northern Corsica - 2003 (about 1200 hectares).

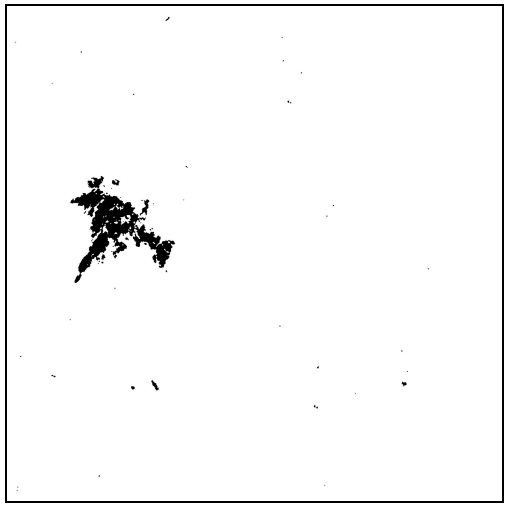
(a) SPOT 5 image © CNES 2005, Distribution SPOT Image



(b) Ground Truth - projet RISK-EOS © Infoterra - ESA

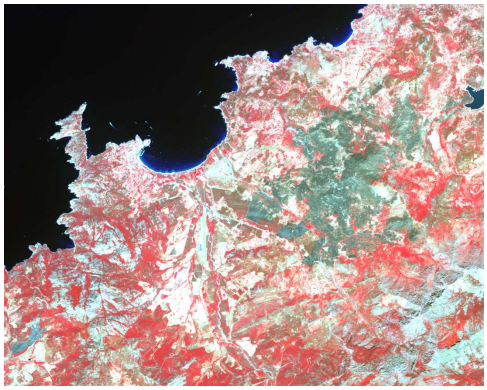


(c) Ground Truth © SDIS2B



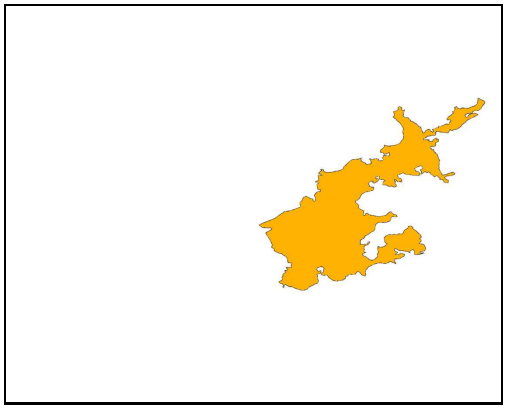
(d) SVM

Figure 16: Eastern Corsica - 2005 (about 1300 hectares).



(a) SPOT 5 image © CNES 2005, Distribution SPOT Image

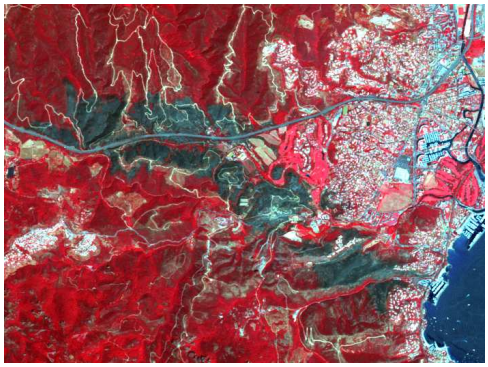
(b) Ground Truth - projet RISK-EOS © Infoterra - ESA



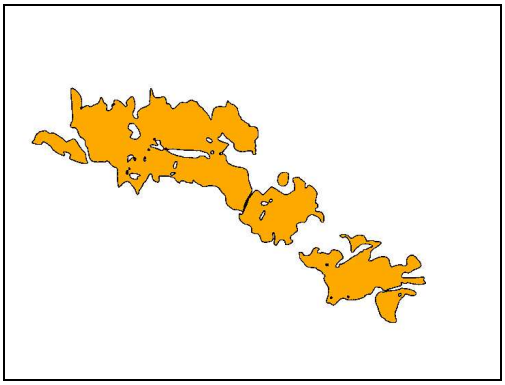
(c) Ground Truth © SDIS2B

(d) SVM

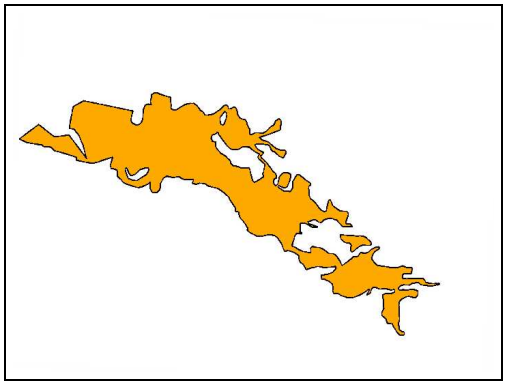
Figure 17: Western Corsica - 2005 (about 2300 hectares).



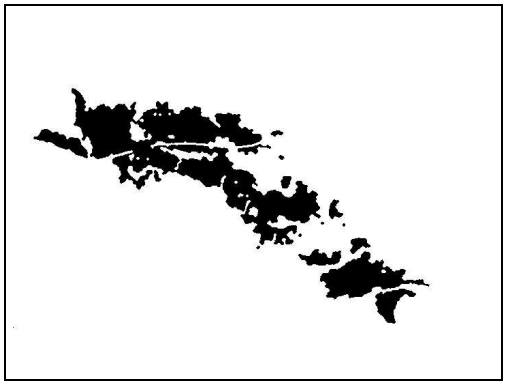
(a) SPOT 5 image © CNES 2007, Distribution SPOT Image



(b) Ground Truth - projet RISK-EOS © Infoterra - ESA

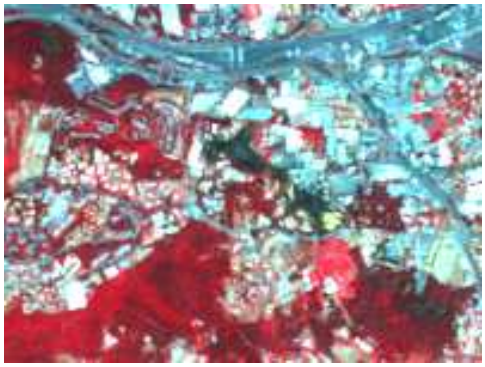


(c) Ground Truth © SDIS06

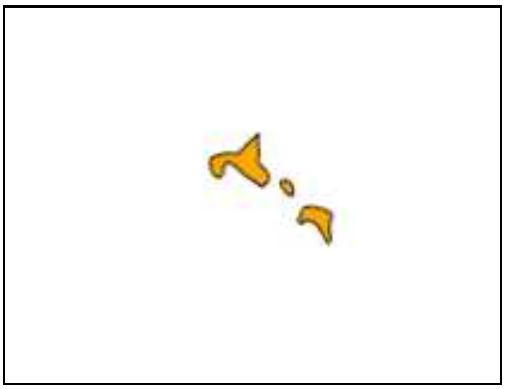


(d) SVM

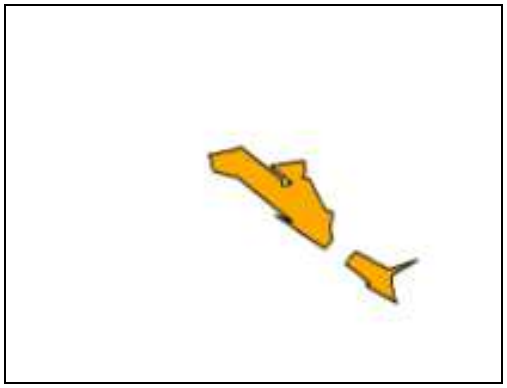
Figure 18: Zoom of Alpes-Maritimes - 2007 (about 450 hectares).



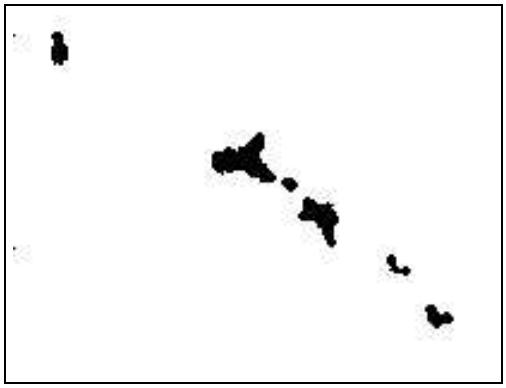
(a) SPOT 5 image © CNES 2007, Distribution SPOT Image



(b) Ground Truth - projet RISK-EOS © Infoterra - ESA

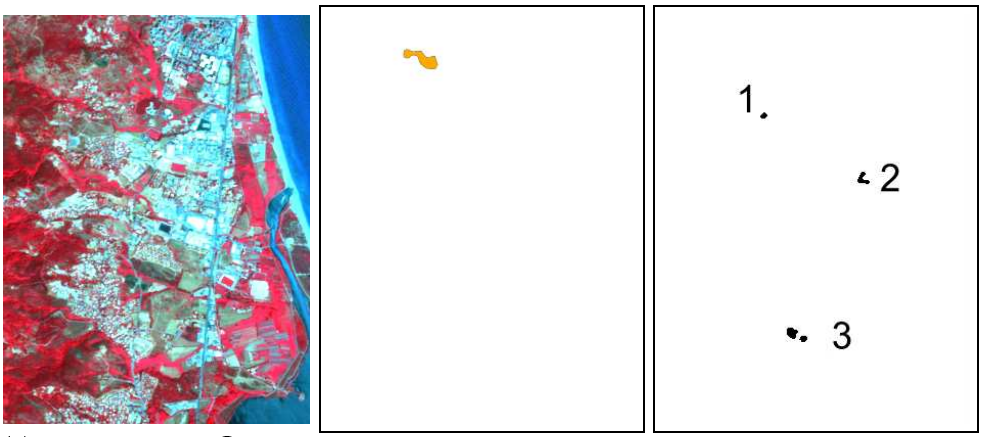


(c) Ground Truth © SDIS06



(d) SVM

Figure 19: Zoom of Alpes-Maritimes - 2007 (about 10 hectares).



(a) SPOT 5 image © CNES 2005, Distribution SPOT Image (b) Ground Truth - projet RISK-EOS © Infoterra - ESA (c) SVM

Figure 20: Zoom of Eastern Corsica -2005 (6 hectares).

3.4.3 Errors analysis

The extracted burnt areas show good agreements with the ground truths. Nevertheless, the differences between the extracted burnt areas and the ground truths can be explained by:

- the errors of the extracted burnt areas due to:

  - the misclassification of features with the same spectral characteristics as burnt areas,
  - the different responses of various types of vegetation after fires. Some areas affected by the fire are not completely burnt but are weakly damaged and thus, can not be discriminated by remote sensing,
  - the mistakes made on the boundaries of the areas: a pixel of the satellite image is considered as burnt if the tree crowns are burnt, a pixel of the ground truth was classified as burnt if it was hit by the fire.

- the errors of the manually made ground truths:

  - the boundaries areas of the ground truths were approximately delineated with ground estimate using GPS (the steeper the slopes in the forest, the more difficult the delineation of the fire scar),
  - the great spatial discontinuity of fires which implies small patches of unburnt vegetation inside the burnt areas.

- other possible errors:

  - the time lag of several months between the fire extinction and the acquisition of the images. The spectral characteristics of the burnt areas could have changed during this period. Even if the revisit frequency of SPOT 5 is about 3 days, some areas are not always well covered due to SPOT 5 oblique viewing and programming capacities and/or we have to choose an image acquired later due to the presence of clouds,
  - the error due to the difference of resolutions between the ground truth and the satellite image,
  - the registration between the ground truth and the satellite image.

3.4.4 Use of vegetation indices

Up to know, the classified data consisted in the spectral signature of the image, i.e. a 3-dimension vector (4-dimension in case of a 4-band image with MIR band).

We also tried to improve the classification by adding vegetation indices (NDVI or NDII) to the data leading to a 4-dimension vector (or respectively to a 5-dimension vector). Note

that these data are normalized to ensure the same scale between the dimensions. When running the SVM algorithm with this new vector, the computing time becomes obviously higher; nevertheless, the classification accuracy remains the same, which means that the addition of vegetation indices in the proposed method does not help to improve it.

4 Conclusion

This research report investigates the burnt area mapping problem using a single post-fire high resolution SPOT 5 image. To assess burnt areas, several techniques were used: the K-means, the K nearest neighbours, the SVM and a proposed approach combining K-means and SVM. This former technique outperforms the K-means and the K-nearest neighbours algorithms.

For future works, we will investigate some methods concerning the automatic hyperparameter estimation. We will also try to improve our results by proposing a new spatial and spectral approach using the texture of the image.

Acknowledgements

The authors would like to thanks the French Space Agency (CNES), via the ISIS program, Infoterra France - ESA, the ONF-AM, the SDIS2B, and the SDIS06 for providing satellite images and ground truths. We also thank Commandant Poppi (fire-brigade member and director of the Cartography service, SDIS83, Draguignan) and Dr. Herve Yesou (project manager, SERTIT-Strasbourg) for interesting discussions. This research has been partly funded by SILOGIC.

References

[1] Chuvieco E., Martin M.P., and Palacios A., "Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination," *International Journal of Remote Sensing*, vol. 23, no. 23, pp. 5103–5110, 2002.

[2] Chuvieco E., *Wildland Fire Danger Estimation and Mapping: The role of Remote Sensing data*, vol. 4 of *Remote Sensing*, World Scientific, 2003.

[3] Sunar F. and Ozkan C., "Forest fire analysis with remote sensing data," *International Journal of Remote Sensing*, vol. 22, no. 12, pp. 2265–2277, 2001.

[4] Barbosa P.M., San-Miguel Ayanz J., Martinez B., and Schmuck G., "Burnt area mapping in Southern Europe using IRS-WiFS," in *Forest Fire Research & Wildland Fire Safety*, Viegas, Ed., Rotterdam, 2002, Millpress.

[5] Chuvieco E., Ventura G., and Martin M.P., "AVHRR multitemporal compositing techniques for burned land mapping," *International Journal of Remote Sensing*, vol. 26, no. 5, pp. 1013–1018, 2005.

[6] Zammit O., Descombes X., and Zerubia J., "Burnt area mapping using support vector machines," in *International Conference on Forest Fire Research*, Figueira da Foz, Portugal, Nov. 2006.

[7] Zammit O., Descombes X., and Zerubia J., "Assessment of different classification algorithms for burnt land discrimination," in *International Geoscience And Remote Sensing Symposium (IGARSS)*, Barcelona, Spain, July 2007.

[8] Vapnik V., *Statistical Learning Theory*, John Wiley and sons, inc., 1998.

[9] Roli F. and Fumera G., "Support vector machines for remote-sensing image classification," in *Image and Signal Processing for Remote Sensing VI*, S. B. Serpico, Ed., Bellingham, 2001, vol. 4170, pp. 160–166, SPIE Proceedings.

[10] Burges C.J.C., "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121–167, 1998.

[11] Vert J.P., "Introduction to support vector machines and applications to computational biology," course notes, Ecole Nationale Supérieure des Mines de Paris, France, 2001.

[12] Schölkopf B., Tsuda K. and Vert J.P., *Kernel Methods in computational biology*, MIT Press, 2004.

[13] Ayat N., Cheriet M., and Suen C., "KMOD — A two parameter SVM kernel for pattern recognition," in *ICPR 2002. Quebec City, Canada*, 2002.

[14] Mercier G. and Lennon M., "Support vector machines for hyperspectral image classification with spectral-based kernels," *Geoscience and Remote Sensing Symposium*, vol. 1, pp. 288–290, 2003.

[15] Schölkopf B., Burges C.J.C. and Smola A.J., *Advances in Kernel Methods: Support Vector Learning*, MIT Press, 1999.

[16] Lasaponara R. and Lanorte A., "Spectral analysis of burned areas observed in the Italian Peninsula by using SPOT-Vegetation data," in *Proc. of the 2nd International Vegetation User Conference*, Antwerp, Belgium, 2004.

[17] Mitra P., Shankar B.U., and Pal S.K., "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067–1074, 2004.

[18] Li S.Z., *Markov random field modeling in computer vision*, Springer-Verlag, 1995.

[19] Winkler G., *Image analysis, random fields and dynamic Monte Carlo methods: a mathematical introduction*, Springer, 1995.

[20] Soille P., *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, 1999.

[21] Schowengerdt R.A., *Remote sensing: models and methods for image processing. Second edition*, Academic Press, 1997.

Unité de recherche INRIA Sophia Antipolis
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes
4, rue Jacques Monod - 91893 ORSAY Cedex (France)
Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399