16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

COMBINING ONE-CLASS SUPPORT VECTOR MACHINES AND HYSTERESIS THRESHOLDING: APPLICATION TO BURNT AREA MAPPING

Olivier Zammit, Xavier Descombes, and Josiane Zerubia

Ariana research group, INRIA-I3S
2004, route des Lucioles, 06902 Sophia Antipolis Cedex, France
phone: + (33) 4 92 38 75 66, fax: + (33) 4 92 38 76 43, email: Firstname.Lastname@sophia.inria.fr
web: www.inria.fr/ariana

ABSTRACT

In this paper, we focus on burnt area mapping using a single post-fire high resolution satellite image. Concerning image classification problems, Support Vector Machines (SVM) have shown great performances. They learn how to distinguish two classes by finding the optimal hyperplane which maximizes the distance between the hyperplane and the training examples. In this paper, we propose to use the One-Class SVM algorithm, an extension of the original two-class SVM which uses only the positive examples in training and testing. This classification algorithm is then followed by a hysteresis thresholding to enhance the image segmentation. To validate the efficiency of the proposed approach, it is tested on high resolution satellite images and the results are compared to the ground truths.

1. INTRODUCTION

Forest Fire is an important environmental issue in forest ecosystems. Each year, thousands of hectares are burnt in the Mediterranean countries, specially during the summer season. Even if forest fires can be beneficial, they endanger the forest biodiversity by changing the biomass stocks and the fauna. Moreover, they damage the soil fertility, the water quality, the hydrological cycles and can also lead to soil erosion.

Assessment of the damage caused by a forest fire plays an important role after the fire extinction. Indeed, an accurate detection provides crucial information for forest offices to plan the restoration and rehabilitation programs. It also helps the fire brigades to better locate the burnt areas, to understand the fire propagation and then, to preempt fight against future fires in the area.

Remote sensing is a valuable tool to assess burnt areas [1]. Several studies have shown the great potential of remote sensing concerning burnt area mapping since it provides spectral information on a large surface, even for remote areas [2]. Most of these studies are based on change detection between multi-temporal series of coarse resolution satellite images, taken before and after the fire [3].

The main interest of our work is to assess the burnt areas from a single post-fire high resolution image [4], [5]. Previously, the discrimination between burnt and unburnt areas was based on the Support Vectors Machines (SVM) algorithm [6], which is a supervised learning technique. This technique provides good results in several contexts such as handwritten digit recognition, pattern recognition or biomedical imagery and has recently been applied to remote sensing classification problem [7]. Moreover, it has proved to outperform classical algorithms such as the K-nearest neighbours or the maximum likelihood.

Given a training set composed of positive and negative pixels (respectively burnt and unburnt pixels), it aims at finding the surface which better separates the two sets. The training set selection represents an important step since it defines the separating surface. The SVM algorithm regards the classification problem as a two-class problem, with equal treatments w.r.t. positive and negative samples. Concerning the problem of discriminating burnt from unburnt areas, it is reasonable to assume that the burnt pixels have similar spectral characteristics, while the ones for unburnt pixels are completely different as they belong to different classes (forest, water, urban areas, roads, fields,...).

In this paper, we use the One-Class SVM (OC-SVM) algorithm, an extension of the original two-class SVM which only uses the positive examples to classify unlabelled pixels. Thus, this avoids selecting the negative training set and reduces the computational cost needed to perform learning and classification. The discrimination between burnt and unburnt pixels is performed from the spectral information of the pixels. Moreover, we propose to add a spatial information by using a hysteresis thresholding [8], which is a well-known technique for edge detection.

This paper is organized as follows: section 2 presents the proposed approach, in section 3 the obtained results are compared to official ground truths and to OC-SVM. Finally, a conclusion is presented in section 4.

2. THE PROPOSED APPROACH

2.1 Support Vector Machines

Consider a training set $\{(\overrightarrow{x_i}, y_i)\}_{i\in<1,N>}$ where $\overrightarrow{x_i} \in R^m$ and $y_i = \pm 1$. Let ϕ be a mapping ($\phi : R^m \rightarrow F$) and K the corresponding kernel: $K(\overrightarrow{x_i}, \overrightarrow{x_j}) = \phi(\overrightarrow{x_i}).\phi(\overrightarrow{x_j})$.

The goal of a classifier is to find the separating hyperplane defined by: $\overrightarrow{w}.\phi(\overrightarrow{x}) + b$ where $(\overrightarrow{w}, b)$ are the parameters of the hyperplane (respectively a vector normal to the hyperplane and the bias).

The final classifier is given by the position w.r.t. the hyperplane:

$$f(\overrightarrow{x}) = sign(\overrightarrow{w}.\phi(\overrightarrow{x}) + b)$$

SVM aim at maximizing the distance between the Optimal Separating Hyperplane (OSH) and the training examples, while minimizing the classification mistakes for the training set (see Fig. 1) [9].

Thus, for most of the training examples, we must have:

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

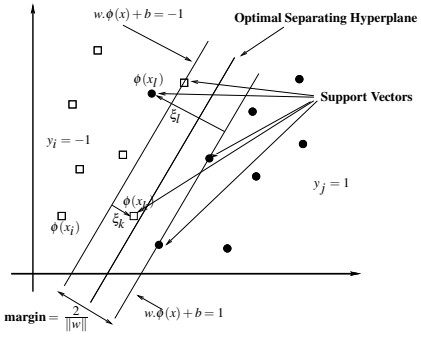


Figure 1: SVM.

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 $\frac{2}{||w||}$. Then, most of the training samples must be on the good side of the tube defined by the equations $\overrightarrow{w}.\overrightarrow{x}+b=-1$ and $\overrightarrow{w}.\overrightarrow{x}+b=+1$.

The formula to be minimized is:

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

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

where C defines the trade-off between a large margin and a small training set error and ξ_i are slack variables introduced to allow some training error while increasing the margin.

By introducing the Lagrange multipliers, the quadratic problem is equivalent to:

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

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

The solution $\overrightarrow{w}$ is then given by: $\overrightarrow{w}=\sum_{i=1}^{N}\lambda_i y_i\phi(\overrightarrow{x_i})$ and b is obtained from the constraints: $\lambda_i.[y_i(\overrightarrow{w}.\phi(\overrightarrow{x_i})+b)-1]=0$

Finally, we have: $f(\overrightarrow{x})=sign(\sum_{i=1}^{N}\lambda_i y_i K(\overrightarrow{x_i},\overrightarrow{x})+b)$. Then, the classification of unlabelled data $\overrightarrow{x}$ is inferred from the training examples such that λ_i is not null. These vectors are called Support Vectors (SV). Moreover, for the SV such that $0<\lambda_i<C$, these vectors are well classified: $f(\overrightarrow{x_i})=sign(y_i)=\pm 1$ and lie at a distance equal to 1 from the OSH.

Notice that the mapping ϕ does not directly interfere with the function f, on the contrary of the kernel K. Then, given a kernel K which has to satisfy Mercer condition [10], we do not have to compute the mapping ϕ. In this study, we choose the well-known Gaussian kernel:

$$K(\overrightarrow{x_i},\overrightarrow{x_j})=exp(-\frac{\|\overrightarrow{x_i}.\overrightarrow{x_j}\|^2}{2\sigma^2})$$

where σ is the variance of the Gaussian.

2.2 One-Class Support Vector Machines

Schölkopf et al. [11] recently extended the SVM methodology to handle training using only positive information. The OC-SVM aim at estimating the support that can include most of the positive samples, i.e, finding a function which is positive in a small region capturing most of the data and negative elsewhere.

This approach is equivalent to find the surface which separates the positive data from the origin at a threshold ρ:

$$f(\overrightarrow{x})=sign(\overrightarrow{w}.\phi(\overrightarrow{x})-\rho)$$

Given l positive training examples $\{\overrightarrow{x_i}\}_{i\in<1,l>}$, the minimization problem is:

$$\min_{(w,\rho,\xi)}\left[\frac{\|\overrightarrow{w}\|^2}{2}-\rho+\frac{1}{\nu.l}\sum_{i=1}^{l}\xi_i\right]$$

$$\text{subject to : } \overrightarrow{w}.\phi(\overrightarrow{x_i})\geq\rho-\xi_i\ ,\ \xi_i\geq 0,\ \forall i\in\langle 1,l\rangle$$

where $\nu\in[0,1]$ defines the trade-off between the margin and the number of training errors.

By introducing the Lagrange multipliers, the quadratic problem is equivalent to:

$$\max_{\lambda} W(\lambda)=-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j K(\overrightarrow{x_i},\overrightarrow{x_j}) \qquad (1)$$

$$\text{subject to : } \sum_{i=1}^{l}\lambda_i=1\ ,\ 0\leq\lambda_i\leq\frac{1}{\nu.l}\ ,\ \forall i\in\langle 1,l\rangle \qquad (2)$$

The solution $\overrightarrow{w}$ is given by: $\overrightarrow{w}=\sum_{i=1}^{l}\lambda_i\phi(\overrightarrow{x_i})$ and b is obtained from the constraints: $\lambda_i.[\overrightarrow{w}.\phi(\overrightarrow{x_i})-\rho]=0$

Finally, we have: $f(\overrightarrow{x})=sign(\sum_{i=1}^{l}\lambda_i K(\overrightarrow{x_i},\overrightarrow{x})-\rho)$. Then, the classification of unlabelled data $\overrightarrow{x}$ is inferred from the training examples such that λ_i is not null. These vectors are called Support Vectors (SV). Moreover, for the SV such that $0<\lambda_i<\frac{1}{\nu.l}$, these vectors lies on the separating surface: $f(\overrightarrow{x_i})=0$. Vectors such that $\lambda_i=0$ are on the good side of the surface: $f(\overrightarrow{x_i})>0$. The ones such that $\lambda_i=\frac{1}{\nu.l}$ are misclassified and on the bad side of the surface: $f(\overrightarrow{x_i})<0$.

2.3 Post-processing

2.3.1 Hysteresis thresholding

The OC-SVM classification is defined by the sign of $\overrightarrow{w}.\phi(\overrightarrow{x})-\rho$, which is equivalent to use a single static threshold value (equal to 0). Nevertheless, this hard binary decision does not achieve good results due to the overlap between the negative and positive classes (some burnt pixels can have spectral characteristics similar to the ones of unburnt pixels). Thus, we propose to use a hysteresis thresholding algorithm [8], which gives well-connected edges while eliminating isolated noisy pixels.

Hysteresis thresholding uses a pair of thresholds, T_{high} and T_{low} with $T_{high}>T_{low}$. The gray-level classification image defined by $\overrightarrow{w}.\phi(\overrightarrow{x})-\rho$ is first segmented by the hard threshold T_{high}, which gives high-confidence pixels (but can also allow some false positives). In our case, these pixels are immediately classified as burnt pixels. The second threshold operation, this time with the weak threshold T_{low}, increases

the size of the burnt clusters previously obtained by the first threshold while new clusters appear. The final segmentation is achieved by choosing, among the clusters selected by the weak threshold, the only ones which are connected to the high-confidence pixels.

To summarize, all the remaining pixels with classification value below T_{low} are totally eliminated (classified as unburnt pixels). Those with value below T_{high} are eliminated unless they are connected to pixels with values above T_{high} (classified as burnt pixels). Hysteresis thresholding enhances the edge of the extracted areas while suppressing some noise.

2.3.2 *Mathematical morphology*

However, some isolated pixels with classification value above T_{high} can be false positives and generate a false positive cluster after the second thresholding. Therefore, an erosion procedure [12] is applied to the first segmentation to eliminate these pixels, considered as noise. This step is understandable since the burnt area size is generally superior to one or two pixels for a high-resolution image.

Moreover, it remains some isolated pixels with classification value below T_{low} inside the positive clusters. Therefore, a closing procedure [12] is applied to the final segmentation. This seems also coherent since one or two isolated pixels can not be completely undamaged inside the burnt areas for a high-resolution image.

3. EXPERIMENTATIONS

3.1 Model selection

The performance of the OC-SVM and the hysteresis threshold algorithms depends on the choice of the parameters (σ and ν for the OC-SVM and T_{high} and T_{low} for the hysteresis thresholding).

A well known way for choosing the SVM parameters is the K-fold cross-validation technique [10]. It consists in dividing the training set in K subsets, training the SVM on K-1 subsets and testing the resulted classification on the remaining subset. The process is repeated K times, one for each subset and the K tests are averaged to produce a mean error. The selected parameters are the ones which provide the smallest mean error.

Concerning OC-SVM, cross-validation technique can not be used since the training set contains only positive samples. Indeed, parameters which generate a classifier which is positive for each point in R^m will produce a mean error equal to 0 but are obviously not relevant.

Notice that Eq. (2) implies that:

$$1 = \sum_{i=1}^{l} \lambda_i \leq \frac{1}{\nu} \; and \; 0 \leq \lambda_i \leq \frac{1}{\nu.l}$$

which means that ν is an upper bound on the fraction of outliers (misclassified training samples) and ν is a lower bound on the fraction of SV. Thus, ν can be set to the estimated fraction of outliers in the training data.

Rätsch et al. [13] proposed a method to choose the OC-SVM parameters by adding some negative points to the training set and selecting the parameters which separate both classes best.

The two thresholds can be set by the user to improve the resulted segmentation. However, they must be in a certain range of values defined by the classifier values of the training set: $\{\overrightarrow{w}.\phi(\overrightarrow{x_i}) - \rho\}_{i \in <1,l>}$.

For instance, Fig. 2 shows the positions of the training set and of the level sets of the OC-SVM classifier, i.e. level sets of $\overrightarrow{w}.\phi(\overrightarrow{x}) - \rho$, in the 2D-case, given a couple of parameters (σ, ν). The red level set represents the OC-SVM separating surface (level set such that $\overrightarrow{w}.\phi(\overrightarrow{x}) - \rho = 0$), the blue crosses and rounds respectively correspond to non-SV (the ones that lie inside the curve) and the SV (the ones that lie on or outside the curve).

Hysteresis thresholding is applied to the OC-SVM classifier. Acceptable values of the T_{high} and T_{low} thresholds respectively correspond to level sets which includes a little less and a little more training vectors than the zero level set (the first curves inside and outside the red curve in Figure 2). The threshold defined by the second level set inside the zero level set is too high since it includes too few training vectors. The one defines by the second level set outside the zero level set is too low since it includes all the training vectors and is too far from some outliers.

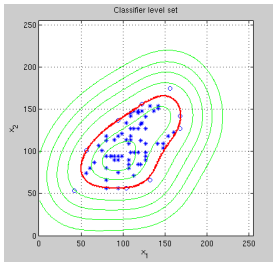


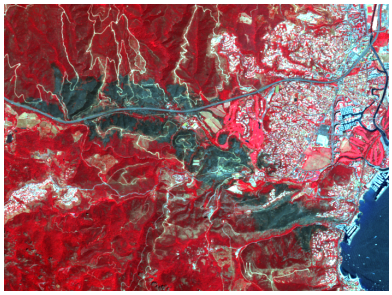
Figure 2: OC-SVM classifier level set

3.2 Tests on high resolution satellite images

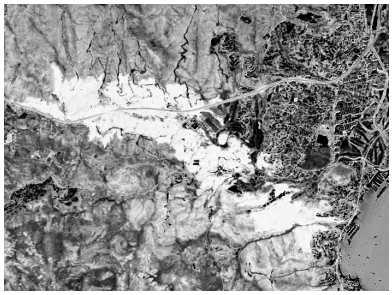
The data used in the experiments are SPOT 5 satellite images representing Southern France areas (the French Riviera and Corsica). Data resolution is 10 meter and they consist in about 2000*2000 pixels with 4 bands (Green, Red, Near Infra-Red and Mid Infra-Red). About 130 training samples were manually selected to form the training set. Only burnt area pixels which are visibly distinct are used as training samples.

The structuring element for the erosion and closing procedures must have a small size in order to avoid erasing true positives and adding false positives. Indeed, due to meteorological conditions and area relief, fire can leave some unburnt patches inside the burnt forest. For our study, we choose a four-connected or eight-connected neighbourhoods. Notice that, for instance, an erosion with a eight-connected structuring element implies removal of clusters whose size is less than 9 pixels, i.e. almost 0.1 hectare for a 10-meter-resolution image.

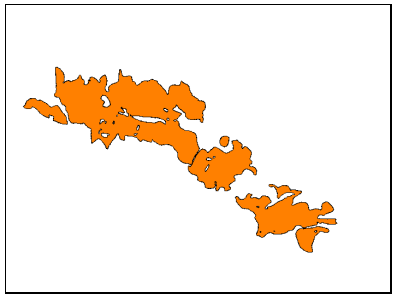
Fig. 3 deals with the area of Cannes-Mandelieu, which burnt in 2007. It shows the satellite SPOT 5 image (3a), the ground truth (3c), the extracted burnt area for the simple OC-SVM classification when no hysteresis is used (3e) and some step of the proposed approach (the OC-SVM classifier, the first segmentation using T_{high}, the final classification using T_{low}). Notice that, for the same reasons as before, opening and closing processes [12] are applied to the simple OC-SVM classification in order to allow a fair comparison be-



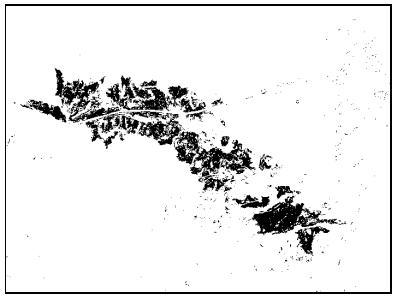
(a) SPOT 5 image © CNES 2007, Distribution SPOT Image



(b) OC-SVM Classifier

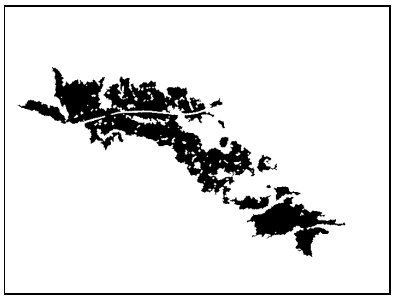


(c) Satellite Estimate - RISK-EOS project ©Infoterra-ESA



(d) OC-SVM classifier thresholding with T_{high}

(e) OC-SVM only



(f) Final extracted burnt areas

Figure 3: Burnt area in Cannes-Mandelieu, France, 2007, size: 450 ha.

tween the two classifications (Fig. 3e-3f). The corresponding burnt area estimate, computed from Landsat satellite images (30-meter-resolution) was given by Infoterra-ESA and constitute our ground truth.

Fig. 3b shows values of the OC-SVM classifier when input data are defined by the four spectral components of pixels. Fig. 3d represents the OC-SVM classifier after thresholding with T_{high}. Then, the isolated pixels are eliminated by the erosion process (considered as noise) and the remaining pixels define seed regions that grow and generate the hysteresis classification after thresholding with T_{low}. Fig. 3f shows the final classification obtained after the closing process.

Fig. 4 show the ROC, i.e. the sensitivity of the proposed approach w.r.t. the two thresholds. The green and red dots correspond respectively to the Figures 3e and 3f.

Fig. 5 deals with the area of Bastia in Corsica, which burnt in 2005. It gives the same subfigures as in Fig. 3.

Several images were used to test the efficiency of the proposed approach but only two are displayed in this paper. When comparing to the ground truths, hysteresis thresholding of the OC-SVM classifier outperforms single thresholding (i.e. simple OC-SVM). Indeed, when combining OC-SVM and hysteresis thresholding, the number of false positives is lower (see Fig. 3) and the edges of the extracted burnt areas are more accurate (see Fig. 5). The main advantages of the proposed method is to use the spatial information to the pixel-based classification technique to improve its results. Moreover, hysteresis thresholding is simple and very fast to use (it takes about 1 second for a 2000*2000 pixel image).

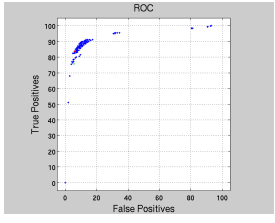
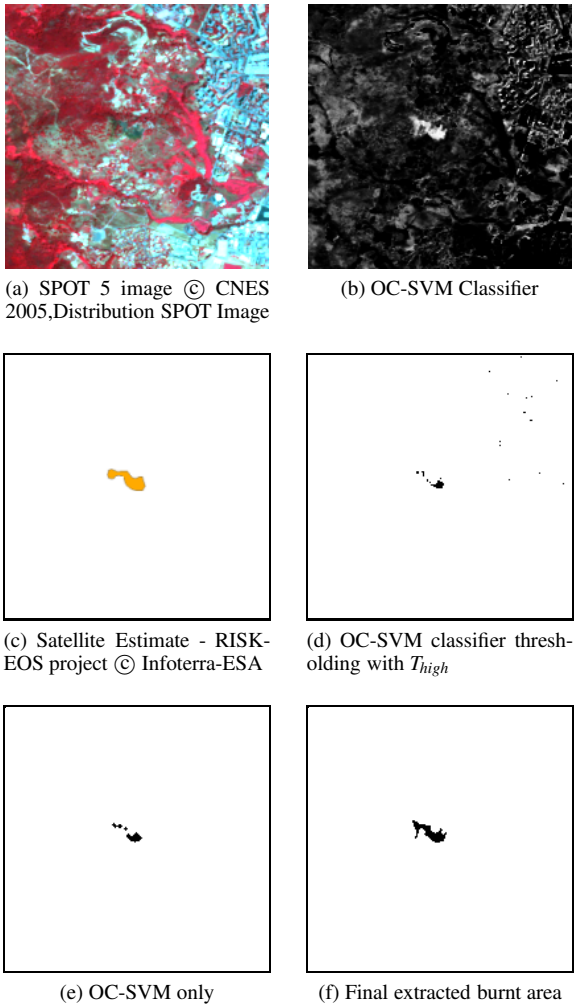


Figure 4: ROC



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

(b) OC-SVM Classifier

(c) Satellite Estimate - RISK-EOS project © Infoterra-ESA

(d) OC-SVM classifier thresholding with T_{high}

(e) OC-SVM only

(f) Final extracted burnt area

Figure 5: Burnt area in Corsica, France, 2005, size: 2 ha.

When using classical SVM algorithm, people can expect better results than the OC-SVM which is obvious since it uses a negative training set. Hysteresis thresholding can also be applied to the SVM classifier in order to improve the classification. Nevertheless, when combining the SVM with a hysteresis thresholding, the resulting accuracy is hardly better compared to the combination of OC-SVM with hysteresis thresholding, while increasing noticeably the computational cost to perform learning and classification.

4. CONCLUSION

In this paper, burnt area mapping using a single post-fire high-resolution image is investigated. The proposed approach is based on an extension of the SVM, the OC-SVM algorithm which only uses positive samples for training and testing. First, for each pixel, OC-SVM give a confidence degree of positivity thanks to pixel spectral values. Then, we use a hysteresis thresholding in order to combine both spectral and spatial information. The obtained results show a better agreement with the ground truths and satellite images than using the OC-SVM only. The hysteresis thresholding technique allows to improve the extracted area edge while decreasing the number of false alarms.

Future work will focus on the automatic selection of the positive training set and the parameters which define the classification.

5. ACKNOWLEDGMENT

The authors would like to thank the French Space Agency (CNES) for providing satellite images via the ISIS program; Infoterra-ESA for providing ground truths. We also thank Commandant Poppi (fire brigade member and director of the cartography service, SDIS83, Draguignan, France) for interesting discussions. This work is supported in part by SILOGIC, Toulouse, France.

REFERENCES

[1] E. Chuvieco, *Wildland Fire Danger Estimation and Mapping: The role of Remote Sensing data*. World Scientific Publishing Company, 2003.

[2] E. Chuvieco, M. P. Martin and A. Palacios, "Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination" *International Journal of Remote Sensing*, vol. 23, pp. 5103–5110, Dec. 2002.

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

[4] O. Zammit, X. Descombes and J. Zerubia, "Burnt area mapping using Support Vector Machines, " in *Proc. International Conference on Forest Fire Research*, Figueira da Foz, Portugal, November 27-30. 2006.

[5] O. Zammit, X. Descombes and J. Zerubia, "Assessment of different classification algorithms for burnt land discrimination," in *Proc. IGARSS*, Barcelona, Spain, July 23-27. 2007.

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

[7] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with Support Vector Machine," *IEEE Trans. Geoscience and Remote Sensing*, vol. 42, pp. 1778–1790, 2004.

[8] J. Canny, "A Computational Approach to Edge Detection," *Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, Nov. 1986.

[9] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, June 1998.

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

[11] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," *Neural Computation*, vol. 13, pp. 1443–1471, July 2001.

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

[13] G. Rätsch, B. Schölkopf, S. Mika, K. R. Müller. "SVM and Boosting: One Class", Technical Report 119. GMD FIRST, Berlin, Germany, Nov. 2000.