

# **Textural kernel for SVM classification in remote sensing: Application to forest fire detection and urban extraction**

Florent Lafarge, Xavier Descombes, Josiane Zerubia

► **To cite this version:**

Florent Lafarge, Xavier Descombes, Josiane Zerubia. Textural kernel for SVM classification in remote sensing: Application to forest fire detection and urban extraction. ICIP - International Conference on Image Processing - 2005, Nov 2005, Genoa, Italy. IEEE, 3, pp.1096-1099, 2005, Image Processing, 2005. ICIP 2005. IEEE International Conference on. <10.1109/ICIP.2005.1530587>. <hal-00814966>

**HAL Id: hal-00814966**

**<https://hal.inria.fr/hal-00814966>**

Submitted on 18 Apr 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# TEXTURAL KERNEL FOR SVM CLASSIFICATION IN REMOTE SENSING : APPLICATION TO FOREST FIRE DETECTION AND URBAN AREA EXTRACTION

*Florent Lafarge, Xavier Descombes and Josiane Zerubia*

Ariana Project, INRIA, 2004 route des Lucioles - BP93, 06902 Sophia Antipolis, Cedex France  
Florent.Lafarge@inria.fr, Xavier.Descombes@inria.fr, Josiane.Zerubia@inria.fr

## ABSTRACT

We present a textural kernel for “Support Vector Machines” classification applied to remote sensing problems. SVMs constitute a method of supervised classification well adapted to deal with data of high dimension, such as images. We introduce kernel functions in order to favor the distinction between our class of interest and the other classes : it gives an information of similarity. In our case this similarity is based on radiometric and textural characteristics. One of the main difficulties is to elaborate textural parameters which are relevant and characterize as well as possible the joint distribution of a set of connected pixels. We apply this method to remote sensing problems : the detection of forest fires and the extraction of urban areas in high resolution images.

## 1. INTRODUCTION

In the mid-nineties, a supervised classification method called “Support Vector Machine” was introduced by Vapnik [1]. This method, derived from the Perceptron problem of Rosenblatt [2], is well adapted to deal with data of high dimension because the algorithm complexity doesn’t depend on the data dimension. That is why SVMs are used in pattern recognition problems, in particular in remote sensing.

In our case, we want to elaborate a general method in order to detect special areas in high resolution images using textural and radiometric characteristics. The kernels are functions introduced to make possible the SVM classification, by specifying the notion of similarity that we are looking for in the pattern recognition problem. In this paper, we propose a new kernel based on a textural information (which depends on the homogeneity of pixel intensity) and a radiometric information (which depends on the hue). One interest of our kernel is to be adjustable by the capability to weight the importance of texture with respect to radiometry by introducing a tuning parameter. It enables to detect various kinds of objects.

We have applied this method to forest fire detection and urban area extraction in SPOT5 satellite images.

## 2. PROPOSED METHOD

### 2.1. Support Vector Machines classification

In this section, we present the SVM classification method. First, we introduce some notations. We call  $E$ , the data space (i.e. the input space) and  $Y$ , the label space (i.e. the output space). Here,  $Y$  is composed of two classes :  $Y = \{-1, 1\}$ . We call  $F$ , a subset of  $E$  which represents the class of interest,  $A$ , the training set (which is a subset of  $E \times Y$ ) and  $\omega$ , the classifier (function of  $E \rightarrow Y$ ).

The theoretical aim of a supervised classification problem is to construct a classifier consistent with the training set. In particular, if we search a linear classifier  $\omega(x) = w \cdot x - \theta$ , the aim is to find the classifier parameters  $(w, \theta)$  which verify

$$\forall (x, y) \in A, \quad y \times (w \cdot x - \theta) > 0 \quad (1)$$

The idea of the SMV is to search the best classifier which gives the biggest margin of security. In fact, as the classifier parameters are defined up to a scale factor, we impose

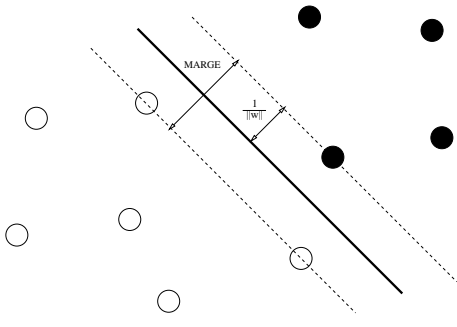
$$\forall (x, y) \in A, \quad y \times (w \cdot x - \theta) > 1 \quad (2)$$

It means we define two hyperplanes parallel to the classifier which are going to lean against the nearest data in order to center as well as possible the classifier between both classes (see figure 1). It is an optimisation problem because we want to find the classifier which maximizes the Euclidian distance between both hyperplanes (called the margin) under the constraint defined by equation (2).

But, in most cases, we cannot find a linear classifier consistent with the training set : the classification problem is not linearly separable. One solution consists of changing the feature space. We introduce a mapping  $\Phi : E \rightarrow \mathcal{E}$  which projects the data in a space of higher dimension where the problem is linearly separable [3].

The dual representation of the optimisation problem is then given by the following formula :

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j (\Phi(x_i) | \Phi(x_j)) \\ \text{under} \quad & \forall i, \alpha_i \geq 0 \\ & \sum_{i=1}^N \alpha_i y_i = 0 \end{aligned} \quad (3)$$



**Fig. 1.** Linear classifier margin

where the  $\alpha_i$  are the Lagrange-Kuhn-Tucker parameters and  $(\cdot, \cdot)$ , the Euclidian scalar product associated with  $\mathcal{E}$ . It is a convex optimisation problem. The optimal solution  $\alpha^*$  gives the optimal classifier parameters  $w^*$  and  $\theta^*$ . Finally, the optimal classifier  $\omega^*$  is given by :

$$\omega^*(x) = \text{sign}_{\mathbb{R}^+} \left( \sum_{i=1}^N \alpha_i^* y_i (\Phi(x_i) | \Phi(x)) - \theta^* \right) \quad (4)$$

Finding the mapping  $\Phi$  is a very difficult problem. Instead of searching for  $\Phi$ , it is easier to search directly for a function  $K : E \times E \rightarrow \mathbb{R}$  defined by  $K(x_i, x_j) = (\Phi(x_i) | \Phi(x_j))$  for  $(x_i, x_j) \in E \times E$  and called a kernel function.

In fact, the knowledge of  $K$  is sufficient to compute the optimal classifier. One of the principal interests of the kernel functions is the faculty to compose several kernels in order to set a more complex kernel which is better adapted to the problem [4]. In the following section, we construct such a kernel.

## 2.2. Textural kernel

The complex patterns we want to detect in satellite images don't make the problem linearly separable. Therefore, it is necessary to introduce a kernel which is able to give relevant informations about the differentiation of classes. This section describes the construction of a kernel based on textural informations.

First, we take a texture parameter derived from a Markovian Gaussian model proposed in [5].

$$P(X_s | X_r, r \in V(s)) = \frac{1}{Z_{V(s)}} \exp \left( -\beta \left( \sum_{r \in V(s)} (X_s - X_r)^2 + \lambda (X_s - \mu)^2 \right) \right) \quad (5)$$

where  $V(s)$  is the neighborhood of the site  $s$ ,  $\lambda$  and  $\beta$  are the parameters of the model and  $\mu$ , the local mean and  $Z_{V(s)}$ ,

the local partition function.

$$P(X_s | X_r, r \in V(s)) \equiv \mathcal{N} \left( \frac{\alpha m_s + \lambda \mu}{\alpha + \lambda}, \frac{1}{2\beta(\alpha + \lambda)} \right) \quad (6)$$

where  $\alpha = \text{card}(V(s))$  and  $m_s = \frac{1}{\alpha} \sum_{r \in V(s)} X_r$ .

This probability follows a normal distribution whose variance will be considered as the texture parameter. It is estimated by the "comet tail" method described in [6].

By defining eight different neighborhoods in eight directions [7], we obtain eight texture parameters  $\sigma_d^2$  for each site of the image.

Now, we are going to construct the kernel. As we are dealing with a Markovian modeling, the data space is composed of the sites of the image and their neighborhoods :  $E = \{(x_s, \{x_r/r \in V(s)\})/s \in S\}$ . We project the data in a new space which includes the texture parameters. For that, we introduce the mapping

$$\begin{aligned} \Phi_1 : E &\rightarrow E \times \Sigma \\ X_s &\rightarrow \Phi_1(X_s) = \begin{pmatrix} X_s \\ \gamma \sigma_d^2 \end{pmatrix} \end{aligned} \quad (7)$$

where  $\Sigma = \{\{\gamma \sigma_d^2/d \in D\}/s \in S\}$ ,  $D$  is the set of the eight directions and  $\gamma$ , a tuning parameter.

Then, we compose it with a Gaussian kernel  $K_g$  in order to obtain a robust kernel. A Gaussian kernel is very useful because it enables to calculate similarity in a space of infinite dimension [4]. Finally, we obtain the following kernel :

$$\begin{aligned} \forall (X_s, X_t) \in E^2 \\ K(X_s, X_t) &= K_g(\Phi_1(X_s), \Phi_1(X_t)) \\ &= \exp \left( -\frac{\|\Phi_1(X_s) - \Phi_1(X_t)\|^2}{2\sigma^2} \right) \end{aligned} \quad (8)$$

where  $\sigma$  is the variance of the Gaussian kernel. The lower the value of  $\sigma$ , the better the learning of an example. But in this case, we must be careful about the over-learning phenomenon.

More explicitly, the kernel can be written as follows :

$$K(X_s, X_t) = e^{\left( \frac{-1}{2\sigma^2} \sum_{(r,q) \in C_{s,t}} (x_r - x_q)^2 + \gamma^2 \sum_{d \in D} (\sigma_{(d),s}^2 - \sigma_{(d),t}^2)^2 \right)} \quad (9)$$

We can see in equation (9) that the kernel is composed of a radiometric term and a textural term which is weighted by the tuning parameter  $\gamma$ . It is a very important parameter : according to the searched pattern, it is necessary to balance the importance of the textural information with respect to the radiometric information, as we will see in the next section.

### 3. RESULTS

We use the soft margin method [8], in order to accept the presence of errors in the classification. This method enables to relax the constraint (2) by introducing variables of relaxation. We violate some constraints in order to stretch the margins of the classifier set.

The obtained classification is then regularized by morphological opening and closing.

The optimisation problem (3) is solved by an ICM algorithm [9], which is modified by adding constraints to the random draw. This algorithm has two parameters : the tuning parameter  $\gamma$  and the variance  $\sigma$ . The calibration of those parameters is made for a range of images.

#### 3.1. Application to forest fire detection

In most cases, we detect forest fires by using ground temperature estimation from Thermal InfraRed images.

But here, we apply our method to panchromatic images from SPOT5 (5m resolution). Our aim is to detect smokes which are emitted by forest fires. The training set is composed of a piece of smoke and a piece of cloud (see figure 2). The presence of a piece of cloud in the training set is necessary to make the distinction between cloud and smoke, due to the similarity of their textures. Figure 3 shows the result. The smoke area is well detected. Contours are accurate. Figure 4 is a result on a second image of the same range, always using the previous training set. There are two small false alarms but the smoke area is well detected. It is important to notice that, on smokeless images, there is practically no over-detection.

In this application, the tuning parameter  $\gamma$  has been chosen so that the radiometry and the texture have the same weight in the similarity criterion.

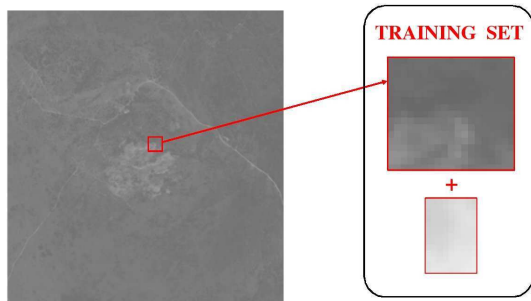


Fig. 2. Training set

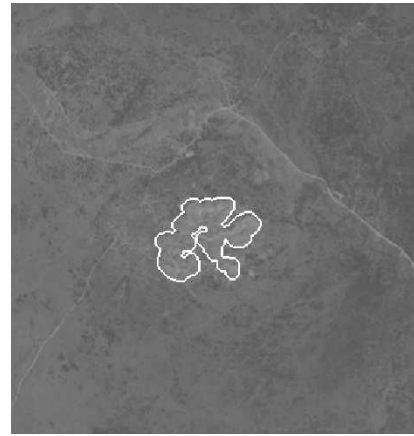


Fig. 3. Result of detection of forest fire smokes on a SPOT5 image of Esterel-Les Maures region (5 meter resolution)

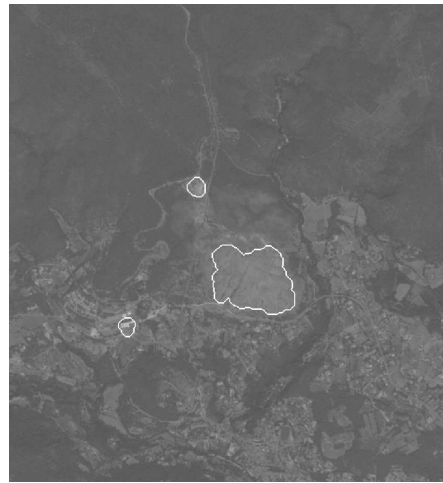
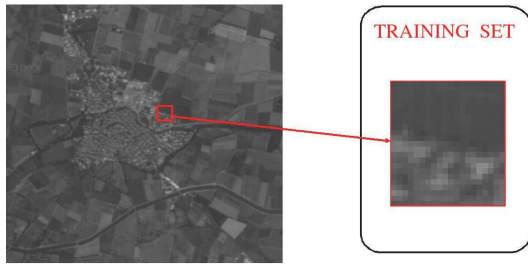


Fig. 4. Result of detection of forest fire smokes on a SPOT5 image of Esterel-Les Maures region (5 meter resolution)

#### 3.2. Application to urban area extraction

The extraction of urban areas is a well known problem. We apply our algorithm to SPOT5 simulations at 5 meters of resolution. The urban areas appear as heterogenous areas without distinction of hues. Therefore, the weight of the textural term in the kernel is increased. More precisely, we take  $\sigma = \gamma \cdot \sigma'$  and we let tend  $\gamma$  to infinity. So, we obtain a kernel which is purely textural.

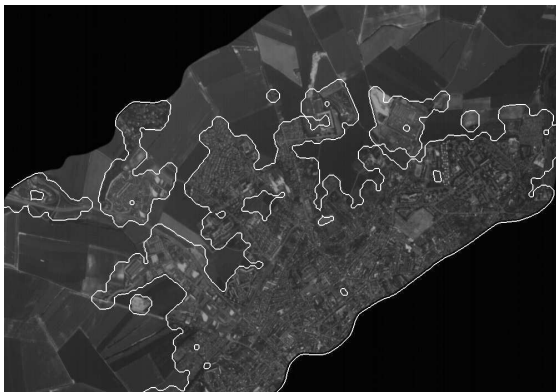
The training set is very simple : as we can see on figure 5, it is a crop of an image of Toreilles (France) composed of a piece of urban areas and a piece of fields. The results (see figures 6 and 7) are satisfactory. There are few false alarms and urban areas are well detected. The separation between urban areas and fields is accurate.



**Fig. 5.** Training set



**Fig. 6.** Result of urban area extraction on a SPOT5 simulation of Toreilles (5 meter resolution)



**Fig. 7.** Result of urban area extraction on SPOT5 simulation of Eppeville (5 meter resolution)

## 4. CONCLUSION

The proposed algorithm gives good results and is able to detect a large range of patterns on high resolution images. This is the main advantage of the proposed method : the adjustable kernel permits to deal with various applications in remote sensing such as forest fire detection or urban area extraction. However, it is necessary to calibrate efficiently the training set for a range of images. In the future, we should work on this problem.

## 5. ACKNOWLEDGEMENTS

The authors thank Alcatel Space Cannes for their partial financial support and CNES and SPOTIMAGE for providing the SPOT5 images of Esterel-Les Maures region via the ISIS program.

## 6. REFERENCES

- [1] V. Vapnik, *The nature of statistical learning theory*, Springer Verlag, 1996.
- [2] F. Rosenblatt, *Principles of Neurodynamics*, Spartan, 1962.
- [3] J. Shawe-Taylor and N. Cristianini, *An Introduction to Support Vector Machines*, The Press Syndicate of the University of Cambridge, 2002.
- [4] M. Pontil and A. Verri, "Properties of support vector machines," Tech. Rep., MIT, 1998.
- [5] X. Descombes, M. Sigelle, and F. Prêteux, "GMRF parameter estimation in a non-stationary framework by a renormalization technique: Application to remote sensing imaging," *IEEE Trans. on Image Processing*, vol. 8(4), pp. 490–503, April 1999.
- [6] X. Descombes, *Champs markoviens en analyse d'images*, Ph.D. thesis, Ecole Nationale Supérieure des Télécommunications de Paris, 1993.
- [7] A. Lorette, X. Descombes, and J. Zerubia, "Texture analysis through a markovian modelling and fuzzy classification: Application to urban area extraction from satellite images," *Int. Journal of Computer Vision*, vol. 36, no. 3, pp. 221–236, 2000.
- [8] M. Samuelides, *Réseaux neuronaux et Apprentissage*, SUPAERO, 2004.
- [9] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Statist. Soc.*, vol. 48, pp. 259–302, 1986.