SVM Classification of High Resolution Urban Satellites Images using Composite Kernels and Haralick Features

Aissam Bekkari, Soufiane Idbraim, Driss Mammass and Mostafa El Yassa
IRF – SIC laboratory,
Faculty of sciences
Agadir -Morocco

Danielle Ducrot
Cesbio, Toulouse- France

***Abstract*—The classification of remotely sensed images knows a large progress taking in consideration the availability of images with different resolutions as well as the abundance of classification's algorithms. A number of works have shown promising results by the fusion of spatial and spectral information using Support vector machines (SVM) which are a group of supervised classification algorithms that have been recently used in the remote sensing field.**

For this purpose we propose a methodology exploiting the properties of Mercer's kernels to construct a family of composite kernels that easily combine multi-spectral features and Haralick texture features as data source.

The proposed approach was tested on common scenes of urban imagery. The three different kernels tested allow a significant improvement of the classification performances and a flexibility to balance between the spatial and spectral information in the classifier. The experimental results indicate an accuracy value of 92.55% which is very promising.

***Index Terms*—SVM, Classification, Composite Kernels, Haralick features, Satellite image, Spectral and spatial information.**

I. INTRODUCTION

With the commercial emergence of the optical satellite images of sub-metric resolution (Ikonos, Quickbird) the realization as well as the regular update of numerical maps with large scales become accessible and increasingly frequent. The classification of such images is similar to that of other image types, it follows the same principle, and it is a method of analysis of data that aims to separate the image into several classes in order to gather the data in homogeneous subsets, which show common characteristics. It aims to assign to each pixel of the image a label which represents a theme in the real study area (e.g. vegetation, water, built, etc) [1].

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [2-4]. Among the most popular and widely used is the maximum likelihood classifier [5]. It is a parametric approach that assumes the class signature in normal distribution. Although this assumption is generally valid, it is invalid for classes consisting of several subclasses or classes having different spectral features [6]. To overcome this problem, some non-parametric classification techniques such as artificial neural networks, decision trees and Support vector machines (SVM) have been recently introduced.

SVM is a group of advanced machine learning algorithms that have seen increased use in land cover studies [7, 8]. One of the theoretical advantages of the SVM over other algorithms (decision trees and neural networks) is that it is designed to search for an optimal solution to a classification problem whereas decision trees and neural networks are designed to find a solution, which may or may not be optimal. This theoretical advantage has been demonstrated in a number studies where SVM generally produced more accurate results than decision trees and neural networks [5, 9]. SVMs have been used recently to map urban areas at different scales with different remotely sensed data. High or medium spatial resolution images (e.g., IKONOS, Quickbird, Landsat (TM)/ (ETM+), SPOT) have been widely employed on urban land use classification for individual cities for ; building extraction, road extraction and other man-made objects extraction [10, 11].

On other hand, the consideration of the spatial aspect in the spectral classification remains very important, for this case, Haralick described methods for measuring texture in gray-scale images, and statistics for quantifying those textures. It is the hypothesis of this research that Haralick's Texture Features and statistics as defined for gray-scale images can be modified to incorporate spectral information, and that these Spectral Texture Features will provide useful information about the image. It is shown that texture features can be used to classify general classes of materials, and that Spectral Texture Features in particular provide a clearer classification of land cover types than purely spectral methods alone.

The proposed method consists of combining spatial and spectral information to obtain a better classification.

doi:10.4304/jetwi.6.1.69-74

We start with the extraction of spectral and spatial information (Haralick texture features) [12]. Then, we apply the SVM classification to the result file.

We exploit the properties of Mercer's kernels to construct a family of composite kernels that easily combine spatial and spectral information. The three different kernels tested demonstrate enhanced classification accuracy as compared to traditional approaches that take into account the spectral information only, and a flexibility to balance between the spatial and spectral information in the classifier

This paper is organized as follows. In the second section, we discuss the extraction of spatial and spectral information especially the Grey-Level Co-occurrence Matrix (GLCM) and Haralick texture features used in experimentations. In section3, we give outlines on the used classifier: Support Vector Machines (SVM). Section4 describe the three different composite kernels used in experimentations. In section5, the results are presented as well as the stating of numerical evaluation. Finally, conclusions are given in section6.

II EXTRACTION OF INFORMATION

2.1 Spectral Information

The most used classification methods for the multispectral data consider especially the spectral dimension. The set of spectral values of each pixel is treated as a vector of attributes which will be directly employed as entry of the classifier. According to Fauvel [13] this allows a good classification based on the spectral signature of each area. However, this does not take in account the spatial information represented by the various structures in the image.

2.2 Spatial Information

Information in a remote sensed image can be deduced based on their textures. A human analyst is able to distinguish man-made features from natural features in an image based on the 'regularity' of the data. Straight lines and regular repetitions of features hint at man-made objects. This spatial information is useful in distinguishing the different field in the remote sensed image.

Many approaches were developed for texture analysis. According to the processing algorithms, three major categories, namely, structural, spectral, and statistical methods are common ways for texture analysis. Grey-Level Co-occurrence Matrix (GLCM) [14] is one of the most widely used methods, which is a powerful technique for measuring texture features; it contains the relative frequencies of the two neighbouring pixels separated by a distance on the image.

Haralick assumed that the texture information is contained in the co-occurrence matrix, and texture features are calculated from it. A large number of textural features have been proposed starting with the original fourteen features described by Haralick et al [15], however only some of these features are in wide use. Wezska et al [16] used four of Haralick features. Conners and Harlow [17] use five features. Conners, Trivedi and Harlow [18] introduced two new features which address a deficiency in the Conners and Harlow set.

We found that the five features used by Conners and Harlow are commonly used because seen that the fourteen are much correlated with each other, and that the five sufficed to give good results in classification [19].

In this work, we have used these five features: homogeneity (E), contrast (C), correlation (Cor), entropy (H) and local homogeneity (LH), and co-occurrence matrices are calculated for four directions: 0°, 45°, 90°and 135° degrees.

Let us recall their definitions:

$$E = \sum_i \sum_j (M(i,j))^2 \tag{1}$$

$$C = \sum_{k=0}^{m-1} k^2 \sum_{|i-j|=k} M(i,j) \tag{2}$$

$$Cor = \frac{1}{\sigma_i \sigma_j} \sum_i \sum_j (i-\mu_i)(j-\mu_j) M(i,j) \tag{3}$$

Where μ_i and σ_i are the horizontal mean and the variance, and μ_j and σ_j are the vertical statistics.

$$H = \sum_i \sum_j M(i,j) \log(M(i,j)) \tag{4}$$

$$LH = \sum_i \sum_j \frac{M(i,j)}{1+(i-j)^2} \tag{5}$$

Each texture measure can create a new band that can be incorporated with spectral features for classification purposes.

III SVM CLASSIFICATION

In this section we briefly describe the general mathematical formulation of SVMs introduced by Vapnik [20, 21]. Starting from the linearly separable case, optimal hyperplanes are introduced. Then, the classification problem is modified to handle non-linearly separable data and a brief description of multiclass strategies is given.

3.1 Linear SVM

For a two-class problem in a *n*-dimensional space R^n, we assume that *l* training samples $x_i \in R^n$, are available with their corresponding labels yi = ±1, S = {(xi, yi) | i ∈ [1, l]}. The SVM method consists of finding the hyperplane that maximizes the margin, i.e., the distance to the closest training data points for both classes [22]. Noting $w \in R^n$ as the normal vector of the hyperplane and $b \in R$ as the bias, the hyperplane H_p is defined as:

$$\langle w, x \rangle + b = 0, \forall x \in H_p \tag{6}$$

Where $\langle w, x \rangle$ is the inner product between w and x. If $x \notin Hp$ then f(x) = $\langle w, x \rangle$ + b is the distance of x to H*p*. The sign of *f* corresponds to decision function *y* = sgn (f(x)).

Finally, the optimal hyperplane has to maximize the margin: $2/\|w\|$. This is equivalent to minimize $\|w\|/2$ and leads to the following quadratic optimization problem:

$$\min\left[\frac{\|w\|^2}{2}\right] \quad (7)$$

$$\text{subject to } y_i(\langle w, x_i\rangle + b) \geq 1 \;\; \forall i \in [1, l]$$

For non-linearly separable data, the optimal parameters (w, b) are found by solving:

$$\min\left[\frac{\|w\|^2}{2} + C\sum_{i=1}^{l}\xi_i\right] \quad (8)$$

$$\text{subject to } y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, \xi_i \geq 0 \;\; \forall i \in [1, l]$$

Where the constant C control the amount of penalty and ξ_i are *slack* variables which are introduced to deal with misclassified samples. This optimization task can be solved through its Lagrangian dual problem:

$$\begin{aligned} &\max_{\alpha} \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j \langle x_i, x_j\rangle \\ &\text{subject to } 0 \leq \alpha_i \leq C \quad \forall i \in [1, l] \\ &\sum_{i=1}^{l}\alpha_i y_i = 0 \end{aligned} \quad (9)$$

Finally:

$$w = \sum_{i=1}^{l}\alpha_i y_i x_i \quad (10)$$

The solution vector is a linear combination of some samples of the training set, whose α_i is non-zero, called Support Vectors. The hyperplane decision function can thus be written as:

$$y_u = \operatorname{sgn}\left(\sum_{i=1}^{l} y_i\alpha_i\langle x_u, x_i\rangle + b\right) \quad (11)$$

Where x_u is an unseen sample.

3.2 Non-Linear SVM

Using the Kernel Method, we can generalize SVMs to non-linear decision functions. With this way, the classification capability is improved. The idea is as follows. Via a non-linear mapping Φ, data are mapped onto a higher dimensional space F:

$$\Phi : R^n \to F \qquad x \mapsto \Phi(x) \quad (12)$$

The SVM algorithm can now be simply considered with the following training samples: $\Phi(S) = \{(\Phi(x_i), y_i) \mid i \in [1, l]\}$. It leads to a new version of the hyperplane decision function where the scalar product is now: $\langle\Phi(x_i), \Phi(x_j)\rangle$. Hopefully, for some kernels function k, the extra computational cost is reduced to:

$$\langle\Phi(x_i), \Phi(x_j)\rangle = k(x_i, x_j) \quad (13)$$

The kernel function k should fulfill Mercers' conditions.

With the use of kernels, it is possible to work implicitly in F while all the computations are done in the input space. The classical kernels used in remote sensing are the polynomial kernel and the Gaussian radial basis function:

$$k_{poly}(x_i, x_j) = [(x_i \cdot x_j) + 1]^p \quad (14)$$

$$k_{gauss}(x_i, x_j) = \exp\left[-\gamma\|x_i - x_j\|^2\right] \quad (15)$$

3.3 Multiclass SVMs

SVMs are designed to solve binary problems where the class labels can only take two values: ±1. For a remote sensing application, several classes are usually of interest. Various approaches have been proposed to address this problem [23]. They usually combine a set of binary classifiers. Two main approaches were originally proposed for a k-classes problem.

- ***One versus the Rest***: k binary classifiers are applied on each class against the others. Each sample is assigned to the class with the maximum output.
- ***Pairwise Classification***: $k(k-1)/2$ binary classifiers are applied on each pair of classes. Each sample is assigned to the class getting the highest number of votes. A vote for a given class is defined as a classifier assigning the pattern to that class.

IV Composite Kernels

In the following section, we present three different kernel approaches for the joint consideration of spectral and textural information for multispectral image classification.

4.1 The Stacked Features Approach

The most commonly adopted approach in multispectral image classification is to exploit the spectral content of a pixel (x_i). However, performance can be improved by including both spectral and spatial information in the classifier. This is usually done by means of the 'stacked' approach, in which feature vectors are built from the concatenation of spectral and spatial features. Note that if the chosen mapping Φ is a transformation of the concatenation $x_i \equiv \{x_{i\text{-spat}}, x_{i\text{-spect}}\}$, then the corresponding 'stacked' kernel matrix is:

$$k_{\{Spect,Spa\}} \equiv k(x_i, x_j) = \langle\Phi(x_i), \Phi(x_j)\rangle \quad (16)$$

which does not include explicit cross relations between $x_{i\text{-spa}}$ and $x_{i\text{-spect}}$.

4.2 The Direct Summation Kernel

A simple composite kernel combining spectral and textural information naturally comes from the concatenation of nonlinear transformations of $x_{i\text{-spat}}$ and $x_{i\text{-spect}}$. Let us assume two nonlinear transformations $\varphi_1(.)$ and $\varphi_2(.)$ into Hilbert spaces H_1 and H_2, respectively. Then, the following transformation can be constructed:

$$\Phi(x_i) = \{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\} \quad (17)$$

and the corresponding dot product can be easily computed as follows:

$$\begin{aligned} k(x_i, x_j) &= \langle\Phi(x_i), \Phi(x_j)\rangle \\ &= \langle\{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\}, \{\varphi_1(x_{j-spect}), \varphi_2(x_{j-spa})\}\rangle \\ &= k_{spect}(x_{i-spect}, x_{j-spect}) + k_{spa}(x_{i-spa}, x_{j-spa}) \end{aligned} \quad (18)$$

4.3 The Weighted Summation Kernel

By exploiting properties of Mercer's kernels, a composite kernel that balances the spatial and spectral content in (19) can also be created, as follows:

$$k(x_i, x_j) = \mu k_{spect}\left(x_{i-spect}, x_{j-spect}\right) + (1-\mu) k_{spa}\left(x_{i-spa}, x_{j-spa}\right) \tag{19}$$

where μ is a positive real-valued free parameter ($0 < \mu < 1$), which is tuned in the training process and constitutes a tradeoff between the spatial and spectral information to classify a given pixel. This composite kernel allows us to introduce a priori knowledge in the classifier by designing specific μ profiles per class, and also allows us to extract some information from the best tuned μ parameter.

Note that solving the minimization problem in all kinds of composite kernels requires the same number of constraints as in the conventional SVM algorithm, and thus no additional computational efforts are induced in the presented approaches.

V Experiments

We developed a two stages classification process: the first one is the extraction of the spatial and spectral features, so we compute Grey Level Co-occurrence Matrix (GLCM) to extract Haralick texture features that we add to spectral information. The second one is the classification stage; with SVM, a supervised kernel learning algorithm widely used.We have selected SVMlight which is an implementation of Support Vector Machines (SVMs) in C language [24] with composite kernels.

To use jointly spatial and spectral information, we used three different kernel approaches presented in section 4; which are the stacked features approach (16), the direct summation kernel (18) and the weighted summation kernel (19).

In the case of the weighted summation kernel, μ was varied in steps of 0.1 in the range [0, 1]. For simplicity and for illustrative purposes, μ was the same for all classes in our experiments. The penalization factor in the SVM was tuned in the range $C = \{10^{-1} \ldots 10^{7}\}$.

We used the Gaussian RBF kernel (15) (with $\sigma = \{10^{-1} \ldots 10^{3}\}$) for the two kernels. k_{spect} uses a spectral information while k_{spa} uses Haralick features.

Concerning data we have used a multispectral satellite image (IKONOS) represented in Fig 1 (a), with size 800 by 600 pixels; at the last we will have, for this image 4 131 individuals (pixels) for learn, 4 952 for validation and 480 000 to classify, divided on six classes (Table 1).

The classification map presented on Fig 1(b), is obtained when the classification is performed using the stacked features approach. When the classification is performed using the direct summation kernel, we obtain the corresponding classification map which is presented on Fig 1(c). A visual analysis of classification maps shows those areas are more homogeneous for the maps obtained using the direct summation kernel.

TABLE 1.
DIFFERENT CLASSES

Class N°	Class name	Train samples	Validation samples
1	Asphalt	1 386	978
2	Green area	480	1 034
3	Tree	196	1 154
4	Soil	813	954
5	Building	920	688
6	Shadow	336	144

The fusion of the spectral and the spatial features using the weighted summation kernel give us the classification map presented on Fig 1(d). The classification map is less noisy and the classification performances are increased globally as well as almost all the classes. It matches well with an urban land cover map in terms of smoothness of the classes; and it also represents more connected classes. Table 2 lists the accuracy estimates for the study area, all models are compared numerically (overall accuracy and kappa coefficient), and table 3, table 4 and table 5 presents respectively the confusion matrix results for SVM classification using the stacked features approach (16), the direct summation kernel (18) and the weighted summation kernel (19).

In conclusion, composite kernels offer excellent performance for the classification of multispectral satellite images by simultaneously exploiting both the spatial and spectral information.

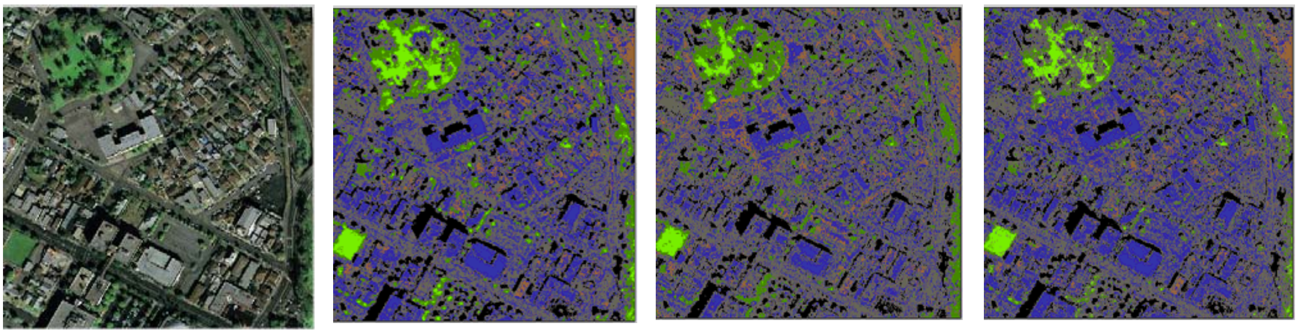
TABLE 2.
OVERALL ACCURACY (%) OF CLASSIFIED IMAGE

Method	OA	Kappa Coefficient
The stacked features approach	92.13%	0.91
The direct summation kernel	92.38 %	0.92
The weighted summation kernel	92.55%	0.92

VI Conclusion

Addressing the classification of high resolution satellite images from urban areas, we have presented three different kernel approaches taking simultaneously the spectral and the spatial information into account (the spectral values and the Haralick features). The weighted summation kernel allows a significant improvement of the classification performances when compared with the two other approaches.

As perspectives, the workflow of this study can be used in other remote sensing application, especially, in rural areas for thematic land cover, and more sophisticated texture techniques to describe the spatial structure of the classes.



(a) (b) (c) (d)

Fig1: (a) Original image, (b) Classification Map obtained using the stacked features approach, (c) Classification Map obtained using the direct summation kernel and (d) Classification Map obtained using the weighted summation kernel.(Asphalt, Green area, Tree, Soil, Building, Shadow)

TABLE 3.
CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE STACKED FEATURES APPROACH.
GLOBAL ACCURACY = 92.13%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**90,12**	1,41	3,92	0	2,63	1,92
Green area	1,13	**94,99**	0	1,08	1,54	1,26
Tree	0,28	1,07	**90,82**	2,5	2,82	2,51
Soil	4,84	0,95	0	**91,87**	2,34	0
Building	3,01	1,16	2,69	2,47	**90,67**	0
Shadow	0,62	0,42	2,57	2,08	0	**94,31**

TABLE 4.
CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE DIRECT SUMMATION KERNEL.
GLOBAL ACCURACY = 92.38%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**89,93**	2,34	0	3,62	2,12	1,99
Green area	1,13	**93,27**	4,71	0	0,53	0,36
Tree	1,18	2,62	**91,22**	1,08	0,6	3,3
Soil	0	0,98	0	**93,95**	5,07	0
Building	6,04	0,46	1,81	0,09	**91,58**	0,02
Shadow	1,72	0,33	2,26	1,26	0,1	**94,33**

TABLE 5
. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE WEIGHTED SUMMATION KERNEL.
GLOBAL ACCURACY = 92.55%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	**89,36**	2,04	1,92	1,5	3,32	1,86
Green area	5,13	**92,21**	0	1,03	1,54	0,09
Tree	1,18	1,52	**93,15**	1,92	0,03	2,2
Soil	1,75	1,13	0,64	**93,04**	3,44	0
Building	1,96	2,78	2,72	0,87	**91,67**	0
Shadow	0,62	0,32	1,57	1,64	0	**95,85**

VII ACKNOWLEDGMENTS

This work was funded by CNRST Morocco and CNRS France Grant under “Convention CNRST CRNS” program SPI09/11.

REFERENCES

[1] Samson C. (2000) "Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles" : Thesis of doctorate, university of Nice-Sophia Antipolis.

[2] Townshend, J.R.G., (1992) "Land cover". International Journal of Remote Sensing 13:1319–1328.
[3] Hall, F.G., Townshend, J.R., Engman, E.T., (1995) "Status of remote sensing algorithms for estimation of land surface state parameters." Remote Sensing of Environment 51:138–156.
[4] Lu, D.,Weng, Q.,(2007) "A survey of image classification methods and techniques for improving classification performance. " International Journal of Remote Sensing 28:823–870.
[5] Huang, C., Davis, L.S., and Townshed, J.R.G.,(2002) "An assessment of support vector machines for land cover classification. "International Journal of Remote Sensing 23:725–749.
[6] Kavzoglu, T., Reis, S.,(2008) "Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels." GIScience and Remote Sensing 45:330–342.
[7] Pal, M., and Mather, P. M. (2005). "Support vector machines for classification in remote sensing." International Journal of Remote Sensing, 26:1007−1011.
[8] Zhu, G., and Blumberg, D. G. (2002). "Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel." Remote Sensing of Environment, 80:233-240.
[9] Scholkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., et al. (1997). "Comparing support vector machines with gaussian kernels to radial basis function classifiers." IEEE Transactions on Signal Processing, 45:2758−2765.
[10] Cao X., Chen J., Imura H., Higashi O., (2009) "A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data", Remote Sensing of Environment 113:2205–2209.
[11] Inglada J., (2007) "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features", ISPRS Journal of Photogrammetry & Remote Sensing 62:236–248.
[12] Bekkari A., Idbraim S., Mammass D. and El yassa M. (2011) "Exploiting spectral and space information in classification of high resolution urban satellites images using Haralick features and SVM" IEEE 2ed International Conference on Multimedia Computing and Systems ICMCS'11 , Ouarzazate, Morocco.
[13] Fauvel M., Benediktsson, J. A., Chanussot J. and Sveinsson, J. R., (2007) “Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles” IEEE International Geoscience and Remote Sensing Symposium, IGARSS 07, Barcelona Spain.
[14] Chiu, W. Y., and Couloigner I. (2004)"Evaluation of incorporating texture into wetland mapping from multispectral images" University of Calgary, Department of Geomatics Engineering, Calgary, Canada, EARSeL eProceedings.
[15] Haralick, R.M., Shanmugam K., and Dinstein I., (1973) "Textural Features for Image Classification." IEEE Transactions on Systems Man and Cybernetics.
[16] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. (1976) "A Comparative Study of Texture measures for Terrain Classification." IEEE Transactions on Systems Man and Cybernetics.
[17] Conners, R.W. and Harlow, C.A., (1980) "A Theoretical Comaprison of Texture Algorithms." IEEE Transactions on Pattern Analysis and Machine Intelligence.
[18] Conners, R.W., Trivedi, M.M., and Harlow, C.A., (1984) "Segmentation of a High-Resolution Urban Scene using Texture Operators." Computer Vision, Graphics and Image Processing.
[19] Arvis V.; Debain C.; Berducat M.; Benassi A. (2004) "Generalization of the cooccurrence matrix for colour images: application to colour texture classification" journal Image Analysis and Stereology, 23:63-72.
[20] Chapel L. (2007) " Maintenir la viabilité ou la résilience d'un système : les machines à vecteurs de support pour rompre la malédiction de la dimensionnalité ? " : Thesis of doctorate, university of Blaise Pascal - Clermont II.
[21] Aseervatham S. (2007) " Apprentissage à base de Noyaux Sémantiques pour le traitement de données textuelles " : Thesis of doctorate, university of Paris 13 –Galilée Institut Laboratory of Data processing of Paris Nord.
[22] Bousquet O., (2001) "Introduction au Support Vector Machines (SVM)", Center mathematics applied, polytechnique school of Palaiseau. http://www.math.u-psud.fr/~blanchard/gtsvm/index.html.
[23] Fauvel M., Chanussot J. and Benediktsson J. A. (2006) "A Combined Support Vector Machines Classification Based on Decision Fusion" IEEE International Geoscience and Remote Sensing Symposium, IGARSS 06, Denver, USA.
[24] SVMlight Version: 6.02 (2008) Developed at University of Dortmund, Informatik, AI-Unit Collaborative Research Center on 'Complexity Reduction in Multivariate Data' (SFB475). http://svmlight.joachims.org/