

SVM Classification of Urban High-Resolution Imagery Using Composite Kernels and Contour Information

Aissam Bekkari, Mostafa El yassa, Soufiane Idbraim,
Driss Mammass and Azeddine Elhassouny
IRF – SIC laboratory,
Faculty of sciences
Agadir, Morocco

Danielle Ducrot
Cesbio laboratory,
Toulouse, France

Abstract—The classification of remote sensing images has done great forward taking into account the image's availability with different resolutions, as well as an abundance of very efficient classification 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, however the addition of contour information to both spectral and spatial information still less explored.

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 composite kernel that gives the best results will be used to introduce contour information in the classification process.

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 a global accuracy value of 93.52%, the addition of contour information, described by the Fourier descriptors, Hough transform and Zernike moments, allows increasing the obtained global accuracy by 1.61% which is very promising.

Keywords—SVM; Contour information; Composite Kernels; Haralick feature; Satellite image; Spectral and spatial information; GLCM; Fourier descriptors; Hough transform ; Zernike moments.

I. INTRODUCTION

The rich spectral information available in remotely sensed images allows the possibility to distinguish between spectrally similar materials [1]. However, supervised classification of satellite images (which assumes prior knowledge in the form of class labels for some spectral signatures) is a very challenging task due to the generally unfavourable ratio between the (large) number of spectral bands and the (limited) number of training samples available *a priori*, which results in the Hughes phenomenon.

The application of originally developed methods for the classification of lower dimensional data sets (such as multispectral images) generally provides poor results when

applied to satellite images, particularly in the case of small training sets [2].

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) [3].

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [4-6]. Among the most popular and widely used is the maximum likelihood classifier [7]. 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 that have different spectral features [8]. 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 [9, 10]. 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 of studies where SVM generally produced more accurate results than decision trees and neural networks [7, 11]. 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, building extraction, road extraction and other man-made objects extraction [12, 13].

On the other hand, the consideration of the spatial aspect in classification remains very important. For this case, Haralick has 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.

As well as the contour information is concerned, survey approaches were developed for pattern recognition. The three most used methods are the Fourier descriptors (FD) classically used to shape recognition and template matching; the Hough transform (HT) which has become a standard tool in computer vision field. It allows the detection of lines, circles or ellipses in a traditional way; it can also be extended to the description of more complex object cases. The third method is the Zernike Moments (ZM) used to extract invariant shapes descriptors to some general linear transformations for the images classification.

This work presents the way adopted in our experiments to incorporate contour information into classification process. We have found that the use of this contour information with both spectral and spatial information allows increasing the accuracy obtained using only spectral and spatial information.

The proposed method consists into combining spatial, spectral and contour information to obtain a better classification. So we have started with the extraction of spatial information (Haralick texture features) [14], and the contour information (Fourier descriptors, Hough transform and Zernike moments). Then, we have used these descriptors combined with spectral values as entry of the SVM classifier. We have exploited the properties of Mercer's kernels to construct a family of composite kernels that easily combine spatial and spectral information. The three different composite kernels tested demonstrate enhanced classification accuracy compared to approaches that take into account only the spectral information, and a flexibility to balance between the spatial and spectral information in the classifier.

An extended version of the composite kernel that gives the best results will be used to introduce contour information in the classification process. The result obtained is compared with the same composite kernel using only spectral and spatial information to measure the contribution of contour information in the classification's overall accuracies.

This paper is organized as follows. In the second section, we will discuss the extraction of spectral, spatial and contour information especially the Grey-Level Co-occurrence Matrix (GLCM), Haralick texture features, Hough transform and Zernike moments used in experimentations. In section 3, we will give outlines on the used classifier: Support Vector Machines (SVM). Section 4 will describe the three different composite kernels used in experimentations. In section 5, the experimentations and results would be presented as well as the numerical evaluation. Finally, conclusions and future research lines would be provided in section 6.

II. EXTRACTION OF INFORMATION

A. Spectral Information

The most used classification methods for the remote-sensing data consider especially the spectral dimension. First attempts to analyze urban area used existing methodologies and techniques developed for land remote sensing, based on signal modeling. Each pixel-vector is regarded as a vector of attributes which will be directly employed as an entry of the classifier.

The traditional approach for classifying remote-sensing data may be summed up as: from the original data set, a feature reduction/selection step is performed according to the classes in consideration, and then classification is carried out using these extracted features. In our work, the step of a feature reduction/selection can be skipped considering that we have used multispectral images such as IKONOS, QUICKBIRD.

According to Fauvel [15] 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.

B. 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.

Many researches have been conducted on the use of Gabor filter banks [16] and co-occurrence matrices [17] for the spatial/spectral classification of multispectral data. Other researches have been conducted with mathematical morphology concepts. Palmason et al. [18] and Fauvel et al. [15] suggest an extraction method of morphological profiles. These profiles are computed on the first principal components of hyperspectral images. Plaza [19] uses also mathematical morphology to extract the endmembers of a hyperspectral image. Some other works [20] combine spectral classification with spatial segmentation based on watershed method.

In [21-23], the authors compare different spatial features in unsupervised classification of hyperspectral images; the studies used Gabor filter banks, co-occurrence matrices, Texture spectra and morphological profiles. The results obtained showed that the haralick features extracted from the co-occurrence matrices give the best performance in classification accuracies.

The GLCM method, proposed by Haralick [24, 25], involves two steps to generate spatial features.

First, the spatial information of a digital image is extracted by a co-occurrence matrix calculated on a pixel neighbourhood (pixel window) defined by a moving window of a given size. Such a matrix contains frequencies of any combination of gray levels occurring between pixel pairs separated by a specific distance and angular relationship within the window. The second step is to compute statistics from the gray level co-occurrence matrix to describe the spatial information according to the relative position of the matrix elements.

Even small, a co-occurrence matrix represents a substantial amount of data that is not easy to handle. This is why Haralick uses these matrices to develop a number of spatial indices that are easier to interpret.

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 [25], however only some of these features are in wide use. Wezcka et al [26] used four of Haralick features. Connors and Harlow [27] use five features. Peng Gong and al. [28] show that these features are much correlated with each other. The authors have used the FORTRAN package TEXTRAN for the spatial feature extraction. The analysis was made on the near-infrared band (0.79-0.89/ μm) with a quantization level of 16.

The interpixel distance was kept constant to 1, and the four main orientations were averaged. The window sizes used were 3x3, 5x5, and 7x7 pixels. Preliminary tests made with larger window sizes did not give satisfactory results. Ten texture features were first generated on a 5x5 pixel window. The three less correlated features were then selected to complete the study. The Fig.1. Represents the Correlation Matrix of the 16 Spatial Features.

	ASM	CON	COR	COV	DIS	ETP	HO	IDM	MA	VAR	AVE	ENT	STD	BW	DD	CS
ASM	*	-0.51	-0.34	-0.33	-0.74	-0.90	0.87	0.91	0.96	-0.44	-0.15	-0.83	-0.59	0.00	-0.02	-0.04
CON		*	0.21	0.72	0.92	0.73	-0.75	-0.64	-0.53	0.90	0.17	0.59	0.87	0.02	-0.03	0.06
COR			*	0.60	0.21	0.41	-0.19	-0.17	-0.28	0.47	0.15	0.40	0.55	0.01	-0.37	-0.15
COV				*	0.61	0.54	-0.46	-0.37	-0.33	0.95	0.16	0.45	0.87	0.00	-0.24	-0.06
DIS					*	0.91	-0.94	-0.87	-0.77	0.80	0.19	0.77	0.86	0.01	0.02	0.07
ETP						*	-0.96	-0.94	-0.90	0.67	0.20	0.88	0.81	0.00	-0.03	0.03
HO							*	0.98	0.90	-0.62	-0.18	-0.83	-0.75	0.00	-0.06	-0.08
IDM								*	0.94	-0.52	-0.16	-0.83	-0.65	0.00	-0.08	-0.08
MA									*	-0.45	-0.15	-0.82	-0.58	0.00	-0.04	-0.05
VAR										*	0.18	0.54	0.93	0.01	-0.16	-0.01
AVE											*	0.20	0.21	0.02	-0.11	-0.05
ENT												*	0.72	0.01	-0.03	0.02
STD													*	0.01	-0.18	-0.02
BW														*	0.21	0.12
DD															*	0.41
CS																*

Fig. 1. The Correlation Matrix of the 16 Spatial Features.

In this work, we have chosen the five features used by Connors and Harlow, which are some of the most commonly used spatial measures and the three less correlated (Fig.1.); we have found that these five sufficed to give good results in classification [29].

These five features are: 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 considering a co-occurrence matrix M :

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

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

Where m is the dimension of the co-occurrence matrix M .

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

Where μ_i and σ_i are the horizontal mean and the variance, 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.

C. Contour Information

Fourier descriptors are classical methods to shape recognition and they have grown into a general method to

encode various shape signatures. Previous experiments have used Fourier descriptors to smooth out fine details of a shape. Also, using the portion of Fourier descriptors to reconstruct an image that smooths out the sharp edges and fine details found in the original shape. Filtering an image with Fourier descriptors provides a simple technique of contour smoothing.

Fourier description of an edge is also used for template matching. Since all the Fourier descriptors except the first one do not depend on the location of the edge within the plane, this provides a convenient method of classifying objects using template matching of an object's contour. A set of Fourier descriptors is computed for a known object. Ignoring the first component of the descriptors, the other Fourier descriptors are compared against the Fourier descriptors of unknown objects. The known object, whose Fourier descriptors are the most similar to the unknown object's Fourier descriptors, is the object the unknown object is classified to. They can also be used for calculation of region area, location of centroid, and computation of second-order moments.

On the other hand, in the detection of specific elements, There are algorithms that, so as to identify these basic forms, attempt to follow the contours to finally bind criteria more or less complex to trace the desired shape. Another approach to this problem is to try to accumulate evidences on a particular form existence, such as a line, a circle or an ellipse. It is this approach that has been adopted in the Hough transform. In recent decades, it has become a standard tool in computer vision field. It allows the detection of lines, circles or ellipses in a traditional way. It can also be extended to description of more complex objects cases.

Moreover, the methods of images representation by moments are among the first to have been applied in pattern recognition. The main motivation is to extract invariant shapes descriptors to some general linear transformations for the images classification. Since the initial work of H. Ming-Kuel [30] in 1962 on invariants derived from the image geometric moments, several approaches have been proposed. Most of these defined moments are expressed as radial moments of the image's circular harmonic functions. The image's Zernike Moments (MZ) were introduced by M.R. Teague [31]. He proposed to use complex polynomials of Zernike orthogonal within the unit circle. These methods are distinguished by the used radial kernel form, which is more or less appropriate to the extraction of invariant descriptors to flat similarities.

In the following we will introduce briefly the Fourier descriptors, the Hough transform and Zernike moments used in experiment to describe the contour information.

1) Fourier Descriptors

The Fourier Descriptors (FD) have been frequently used as features for image processing, remote sensing, shape recognition and classification.

The use of FDs for pattern recognition tasks started in the early sixties by Cosgriff [32] and Fritzsche [33]. A set of orthogonal FDs represent each pattern for the purpose of classification. The recognition system was independent of the character size and orientation. Furthermore, FDs were used as features for recognition systems for both handwritten characters

[34] and numerals [35]. Granlund [34] used a small number of lower-order descriptors for the classification system. Those descriptors were insensitive to translation, rotation and dilation. Because of the small computational power available at that time the system could not be examined to give the suitable number of descriptors. The classification system was applied to a small number of characters. Nevertheless the system was able to produce a very good recognition rate of 98%.

Zhan and Roskies [35] started computing the FDs by translating the contour of handwritten numeral into a change of angle curve. A large number of Fourier coefficients are produced. For each coefficient two kinds of FDs are computed, the harmonic amplitude and the phase angle. Those pair of FDs is invariant under translation, rotation and change of size of the original handwritten numeral. All the FDs pairs fully describe the original signature.

Fourier descriptors were also used to describe open curves in an online character recognition system [36]. The one pixel thick strokes were taken online using a tablet. Then twenty FDs were computed and used for classification.

In remote sensing field the FDs were applied to the feature of the regions on the data for geometrical matching of the remote sensing images. It makes possible to monitor natural and artificial changes in land cover precisely.

The discrete Fourier function for a periodic polynomial function $f(t)$ is,

$$F(k) = \frac{1}{N} \sum_{t=0}^{N-1} f(t) \exp(-j2\pi kt / N), \quad (6)$$

Where N is the total number of points along the $f(t)$

$$F(k) = \frac{1}{N} \sum_{t=0}^{N-1} f(t) \{ \cos(2\pi kt / N) + j \sin(2\pi kt / N) \} \quad (7)$$

with $k = 0, 1, 2, \dots, N-1$

The Fourier coefficients are

$$\begin{aligned} a_k &= \frac{2}{N} \sum_{t=1}^{N-1} f(t) \cos\left(\frac{2\pi kt}{N}\right) \\ b_k &= \frac{2}{N} \sum_{t=1}^{N-1} f(t) \sin\left(\frac{2\pi kt}{N}\right) \end{aligned} \quad (8)$$

As said before the commonly used FDs are the harmonic amplitude A_k and the phase angle θ_k of the Fourier coefficients a_k and b_k above,

$$\begin{aligned} A_k &= \sqrt{a_k^2 + b_k^2} \\ \theta_k &= \tan^{-1} \frac{b_k}{a_k} \end{aligned} \quad (9)$$

The harmonic amplitude A_k is a pure shape feature and doesn't contain information about the position or the orientation of the numeral but on the other hand the phase angle θ_k has those two features.

The fixed length feature vector would be

$$\{A_k, \theta_k\}_{k=1}^M, \quad (10)$$

Where M is a fixed integer number.

The original polynomial could be reconstructed from its FDs by using the following equation

$$f(t) \approx A_0 + \sum_{k=1}^{N-1} a_k \cos \frac{2\pi kt}{N} + b_k \sin \frac{2\pi kt}{N} \quad (11)$$

with $t = 0, 1, 2, \dots, N-1$

Where, A_0 is the DC component of the function, and has no effect on the shape description.

2) Hough Transform

The Hough Transform (HT) is considered as a very powerful tool for detecting predefined features (i.e. lines, ellipses...) in images and has been used for more than three decades in the areas of image processing, pattern recognition and computer vision. Its main advantages are its insensitivity to noise and its capability to extract lines even in areas with pixel absence (pixel gaps) [37-39].

The Hough technique is particularly useful for computing a global description of a feature(s) (where the number of solution classes need not to be known *a priori*), given (possibly noisy) local measurements. The motivating idea behind the Hough technique for line detection is that each input measurement (e.g. coordinate point) indicates its contribution to a globally consistent solution (e.g. the physical line which gave rise to that image point).

As a simple example, consider the common problem of fitting a set of line segments to a set of discrete image points (e.g. pixel locations output from an edge detector). Fig.2. shows some possible solutions to this problem. Here the lack of *a priori* knowledge about the number of desired line segments (and the ambiguity about what constitutes a line segment) render this problem under-constrained.

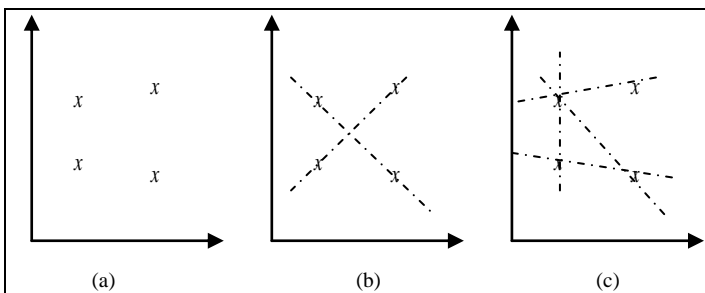


Fig. 2. (a) Coordinate points, when (b) and (c) Possible straight line fittings.

We can analytically describe a line segment in a number of forms. However, a convenient equation for describing a set of lines uses parametric or normal notion as follow:

$$\rho = x \cos \theta + y \sin \theta \quad (12)$$

Where ρ is the length of a normal from the origin to this line and θ is the orientation of ρ with respect to the X-axis. (Fig.3.) For any point (x, y) on this line, ρ and θ are constant.

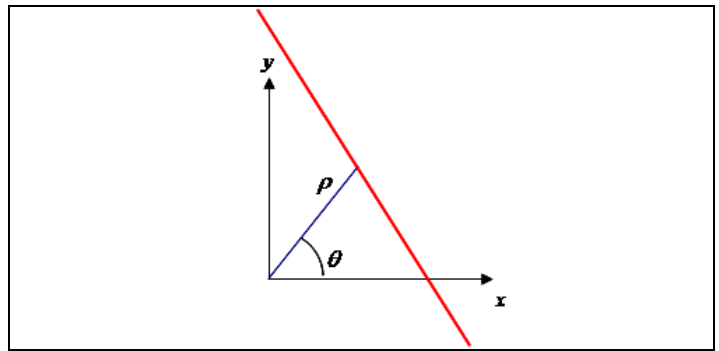


Fig. 3. Parametric description of a straight line (ρ, θ)

In an image analysis context, the coordinates of the point(s) of edge segments (i.e. (x_i, y_i)) in the image are known and therefore serve as constants in the parametric line equation, while ρ and θ are the unknown variables we seek. We plot the possible (ρ, θ) values defined by each (x_i, y_i) points in Cartesian image space map to curves (i.e. sinusoids) in the polar Hough parameter space.

This *point-to-curve* transformation is the Hough transformation for straight lines. When viewed in Hough parameter space, points which are collinear in the cartesian image space become readily apparent as they yield curves which intersect at a common (ρ, θ) point.

The transform is implemented by quantizing the Hough parameter space into finite intervals or *accumulator cells*. As the algorithm runs, each (x_i, y_i) is transformed into a discretized (ρ, θ) curve and the accumulator cells which lie along this curve are incremented. Resulting peaks in the accumulator array represent strong evidence that a corresponding straight line exists in the image.

3) Zernike Moments

The extraction of features from an image by the method of moments is one of the techniques commonly used. It obviously gives the amount of information which is encoded in the image [40]. A moment is an overall description of the distribution of pixels within an image. Each time a given order gives different information of other times on the image [41, 42]. The central moments of order p, q are given by the following expressions:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y)$$

With:
$$m_{pq} = \sum_x \sum_y X^p Y^q I(x, y) \quad (13)$$

$$\bar{x} = \frac{m_{10}}{m_{00}} \quad \text{and} \quad \bar{y} = \frac{m_{01}}{m_{00}}$$

Where $I(x, y)$ is the gray level of the pixel x, y . The central moments are given as following [39, 40]:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \quad (14)$$

The normalized central moments are given by

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \text{with } \gamma = \frac{p+q}{2} + 1 \quad (15)$$

Hu moments are defined as a set of moment invariants [43], but are not orthogonal. The most interesting moments are orthogonal that can be obtained through the Zernike polynomials. The Zernike moments do not change the orientation, the scale and the translation. They remain robust to noise and to minor variations of the forms [44]. There is no redundant information because their bases are orthogonal. An image is best described by a small set of Zernike moments than any other type of moments such as geometric moments, Legendre, rotational or complex moments [45]. The Zernike moments are build using a set of complex polynomials which form a complete orthogonal set on the unit disk. For an image f , the Zernike moments are defined as follows [45]:

$$Z_n = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(\rho, \theta) \quad (16)$$

Where m and n define the order of the moment. Knowing that

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{+jm\theta} \quad (17)$$

Where $R_{nm}(\rho)$ is the radial polynomial Zernike. The latter can be described by:

$$R_{nm}(\rho) = \sum_{s=0}^{n-|m|/2} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (18)$$

Where n and m are integers (their values are even integers). These moments can be used as a tool for comparing two classes by calculating the distance denoted by d between the vectors of Zernike moments of each class. If we are interested in comparing one class to multiple classes, the most similar image corresponds to that which is characterized by a smallest distance d .

III. SVM CLASSIFICATION

In this section we will briefly describe the general mathematical formulation of SVMs introduced by Vapnik [46, 47]. Starting from the linearly separable case in which the optimal hyperplanes are introduced. Then, the classification problem is modified to handle non-linearly separable data. At the end of this section, a brief description of multiclass strategies would be given.

A. 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 $y_i = \pm 1$, $S = \{(x_i, y_i) \mid i \in [1, l]\}$.

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 [48]. 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:

$$\langle w, x \rangle + b = 0, \forall x \in H_p \quad (19)$$

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

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:

$$\min \left[\frac{\|w\|^2}{2} \right] \quad (20)$$

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

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

$$\min \left[\frac{\|w\|^2}{2} + C \sum_{i=1}^l \xi_i \right] \quad (21)$$

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

Where the constant C control the amount of penalty and ξ_i are slack variables which are introduced to deal with misclassified samples (Fig.4.). This optimization task can be solved through its Lagrangian dual problem:

$$\max_{\alpha} \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle \quad (22)$$

subject to $0 \leq \alpha_i \leq C \quad \forall i \in [1, l]$

$$\sum_{i=1}^l \alpha_i y_i = 0$$

Finally:

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (23)$$

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:

$$y_u = \text{sgn} \left(\sum_{i=1}^l y_i \alpha_i \langle x_u, x_i \rangle + b \right) \quad (24)$$

Where x_u is an unseen sample.

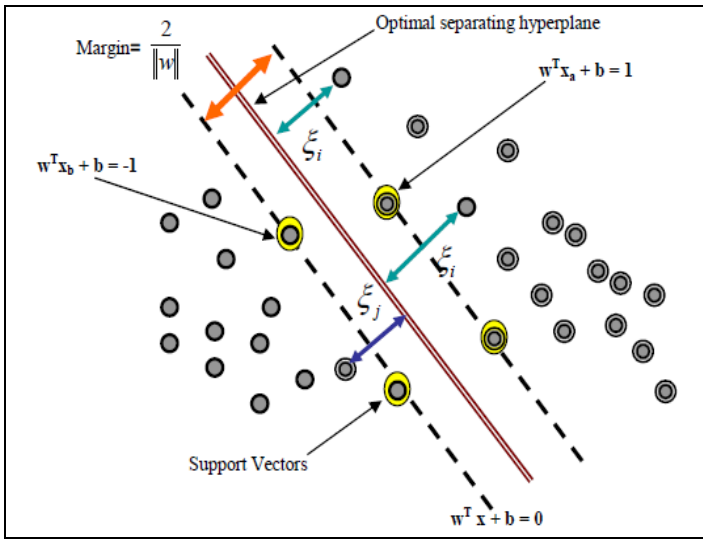


Fig. 4. Classification of a non-linearly separable case by SVMs. There is one non separable feature vector in each class.

B. Non-Linear SVM

Using the Kernel Method, we can generalize SVMs to non-linear decision functions. By this technique, the classification capability is improved. The idea is as following. Via a non-linear mapping Φ , data are mapped onto a higher dimensional space F (Fig.5.):

$$\Phi: R^n \rightarrow F \quad (25)$$

$$x \alpha \Phi(x)$$

The SVM algorithm can now be simply considered with the following training samples: $\Phi(S) = \{(\Phi(x_i), y_i) / i \in [1, l]\}$. 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:

$$\langle \Phi(x_i), \Phi(x_j) \rangle = k(x_i, x_j) \quad (26)$$

The kernel function k should fulfill Mercers' conditions.

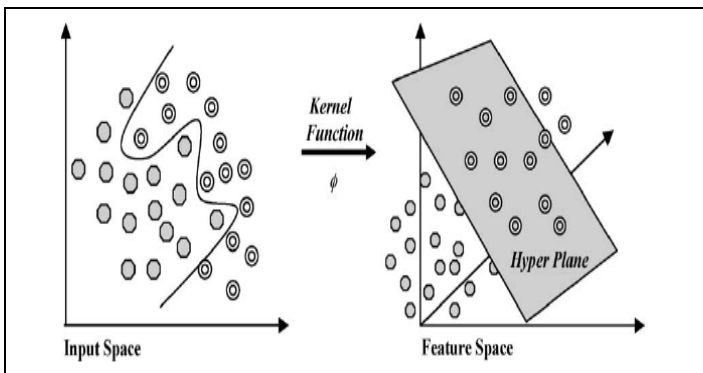


Fig. 5. Mapping the Input Space into a High Dimensional Feature Space with a kernel function

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:

$$k_{poly}(x_i, x_j) = [(x_i \cdot x_j) + 1]^p \quad (27)$$

$$k_{gauss}(x_i, x_j) = \exp[-\gamma \|x_i - x_j\|^2] \quad (28)$$

C. 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 [49]. 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 will be dealing with three different kernel approaches that not only allow joining spectral and textural information for multispectral image classification, but also introducing the contour information by using an extended kernel version [50, 51].

A. 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 $x_i \equiv \{x_{i-spect}, x_{i-spa}\}$, then the corresponding 'stacked' kernel matrix is:

$$k_{\{Spect, Spa\}} \equiv k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle \quad (29)$$

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

Including the contour information is also possible by means of the 'stacked' approach; the feature vectors will be built from the concatenation of spectral, spatial and contour features:

$$x_i \equiv \{x_{i-spect}, x_{i-spa}, x_{i-cont}\}.$$

The corresponding 'stacked' kernel matrix $k_{\{Spect, Spa, Cont\}}$ remains the same in (29).

B. The Direct Summation Kernel

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

$$\Phi(x_i) = \{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\} \quad (30)$$

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

$$\begin{aligned} k(x_i, x_j) &= \langle \Phi(x_i), \Phi(x_j) \rangle \\ &= \langle \{\varphi_1(x_{i-spect}), \varphi_2(x_{i-spa})\}, \{\varphi_1(x_{j-spect}), \varphi_2(x_{j-spa})\} \rangle \\ &= k_{spect}(x_{i-spect}, x_{j-spect}) + k_{spa}(x_{i-spa}, x_{j-spa}) \end{aligned} \quad (31)$$

In the same way, we can exploit the Mercer's properties to generalize this formulation in order to have a summation of multiple kernels:

$$k(x_i, x_j) = \sum_{m=1}^p k_m(x_i^m, x_j^m) \quad (32)$$

So to use spectral, spatial and contour information we take the case of $p=3$, then we will have:

$$k(x_i, x_j) = k_{spect}(x_{i-spect}, x_{j-spect}) + k_{spa}(x_{i-spa}, x_{j-spa}) + k_{Cont}(x_{i-Cont}, x_{j-Cont}) \quad (33)$$

C. The Weighted Summation Kernel

By exploiting properties of Mercer's kernels, a composite kernel that balances the spatial and spectral content in (28) can also be created, as follows:

$$k(x_i, x_j) = \mu k_{spect}(x_{i-spect}, x_{j-spect}) + (1-\mu)k_{spa}(x_{i-spa}, x_{j-spa}) \quad (34)$$

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.

A generalization of the weighted summation to multiple kernels is possible by using "Linear combination methods", and we can linearly parameterize the combination function:

$$k_\mu(x_i, x_j) = \sum_{m=1}^p \mu_m k_m(x_i^m, x_j^m) \quad (35)$$

Where μ denotes the kernel weights. Different versions of this approach differ in the way they put restrictions on μ : the linear sum (i.e., $\mu \in \mathbb{R}^p$), the conic sum (i.e., $\mu \in \mathbb{R}_+^p$), or the convex sum (i.e., $\mu \in \mathbb{R}_+^p$ and $\sum_m \mu_m = 1$). As can be seen, the conic sum is a special case of the linear sum and the convex sum is a special case of the conic sum. The conic and convex sums have two advantages over the linear sum in terms of interpretability.

First, when we have positive kernel weights, we can extract the relative importance of the combined kernels by looking at them. Second, when we restrict the kernel weights to be nonnegative, this corresponds to scaling the feature spaces and using the concatenation of them as the combined feature representation:

$$\Phi_\mu(x) = \begin{pmatrix} \sqrt{\mu_1} \Phi_1(x^1) \\ \sqrt{\mu_2} \Phi_2(x^2) \\ \vdots \\ \sqrt{\mu_p} \Phi_p(x^p) \end{pmatrix} \quad (36)$$

And the dot product in the combined feature space gives the combined kernel:

$$\begin{aligned} \langle \Phi_\mu(x_i), \Phi_\mu(x_j) \rangle &= \begin{pmatrix} \sqrt{\mu_1} \Phi_1(x_i^1) \\ \sqrt{\mu_2} \Phi_2(x_i^2) \\ \vdots \\ \sqrt{\mu_p} \Phi_p(x_i^p) \end{pmatrix}^T \begin{pmatrix} \sqrt{\mu_1} \Phi_1(x_j^1) \\ \sqrt{\mu_2} \Phi_2(x_j^2) \\ \vdots \\ \sqrt{\mu_p} \Phi_p(x_j^p) \end{pmatrix} \\ &= \sum_{m=1}^p \mu_m k_m(x_i^m, x_j^m) \end{aligned} \quad (37)$$

The combination parameters can also be restricted using extra constraints, such as the l_p -norm on the kernel weights or trace restriction on the combined kernel matrix, in addition to their domain definitions. For example, the l_1 -norm promotes sparsity on the kernel level, which can be interpreted as feature selection when the kernels use different feature subsets.

So to use spectral, spatial and contour information we take the case of $p=3$, then we will have:

$$\begin{aligned} k(x_i, x_j) &= \mu_1 k_{spect}(x_{i-spect}, x_{j-spect}) + \\ &\quad \mu_2 k_{spa}(x_{i-spa}, x_{j-spa}) + \mu_3 k_{Cont}(x_{i-Cont}, x_{j-Cont}) \end{aligned} \quad (38)$$

with $\mu_m \in \mathbb{R}_+$ and $\sum_{m=1}^3 \mu_m = 1$

D. The Computational Complexity

The computational complexity of a multiple kernel learning (MKL) algorithm mainly depends on its training method (i.e., whether it is one-step or two-step) and the computational complexity of its base learner.

One-step methods using fixed rules and heuristics generally do not spend much time to find the combination function parameters, and the overall complexity is determined by the complexity of the base learner to a large extent. One-step methods that use optimization approaches to learn combination parameters have high computational complexity, due to the fact that they are generally modeled as a semi-definite programming (SDP) problem, a quadratically constrained quadratic programming (QCQP) problem, or a second-order cone programming (SOCP) problem. These problems are much harder to solve than a quadratic programming (QP) problem used in the case of the canonical SVM.

Two-step methods update the combination function parameters and the base learner parameters in an alternating manner. The combination function parameters are generally updated by solving an optimization problem or using a closed-form update rule. Updating the base learner parameters usually requires training a kernel-based learner using the combined kernel. For example, they can be modeled as a semi-infinite linear programming (SILP) problem, which uses a generic linear programming (LP) solver and a canonical SVM solver in the inner loop.

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. EXPERIMENTAL RESULTS

In this section, we are going to evaluate the proposed approach by using two high resolution satellite images with different resolutions representing the scene of urban areas.

A. Data

The first image used in classification is a subset of high resolution QUICKBIRD satellite image, with a high spatial resolution of 2.4 m per pixel. It represents urban scene areas. We dispose of four spectral bands: blue, green, red and near infrared. We can see in Fig.7. (a) a representation of this subset.

The second image is a subset of high resolution IKONOS satellite image. It has also four spectral bands: red, blue, green and near infrared, with a high spatial resolution of 4.1 m per pixel. This subset of the image is represented in Fig.8. (a).

We will have two files containing the extracted features for each image, "TrainFile.dat" and "TestFile.dat" respectively for learning and for classification, and divided on six classes as described in the following Table I.

B. Comparing Composite Kernels

Our experiments are divided on two stages (Fig.6. and Fig.9.). The first one concerns the studies of composite kernels proposed in section 4 using only spectral and spatial information. In the second stage we will use an extended version of the composite kernel that gave the best performance in the first stage, to introduce contour information in addition to spectral and spatial information.

TABLE I. DIFFERENTS CLASSES

Class N°	Class name	Train samples	
		Image 1	Image 2
1	Asphalt	1 592	1 386
2	Green area	2 252	480
3	Tree	880	196
4	Soil	176	813
5	Building	4 217	920
6	Shadow	1 280	336
Total		10 397	4 131

So as we can see in Fig.6., that represents the first experience, we have developed a two step 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 have added to spectral information. The second step is the SVM classification; a supervised kernel learning algorithm widely used. We have selected SVMlight with composite kernels, which is an implementation of Support Vector Machines (SVMs) in C language [52].

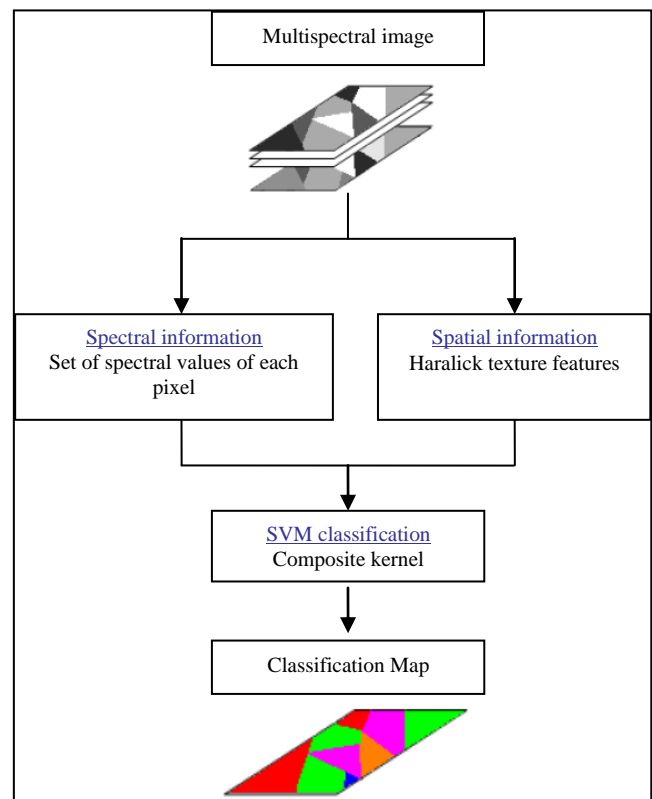


Fig. 6. A representative illustration of the first stage of the proposed workflow

To join spatial and spectral information, we have used three different kernel approaches as presented in section 4; named the stacked features approach in (29), the direct summation kernel in (31) and the weighted summation kernel in (34).

In the case of the weighted summation kernel, μ was varied with a step 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 $C = \{10^{-1} \dots 10^7\}$.

We have used the Gaussian RBF kernel (28) (with $\sigma = \{10^{-1} \dots 10^3\}$) for the two kernels. k_{spect} uses a spectral information while k_{spa} uses Haralick features.

The classification map presented on (b) in Fig.7. and Fig.8., is obtained when the classification is performed using the stacked features approach (29). When the classification is performed using the direct summation kernel (31), we obtain the corresponding classification map which is presented on (c) in Fig.7. and Fig.8.. A visual analysis of classification maps shows those areas more homogeneous for the maps obtained using the direct summation kernel than those obtained by using the stacked features approach.

The fusion of the spectral and the spatial features using the weighted summation kernel give us the classification map

presented on (d) in Fig.7. and Fig.8.. We can see that the classes are more connected and also we have got less misclassified pixels in the result compared to the other approaches.

Table II lists the accuracy estimates and kappa coefficient of the classification results, all models are compared numerically (overall accuracy, kappa coefficient).

Table III and Table IV presents respectively the confusion matrix results for SVM classification using the weighted summation kernel (34) based on spectral and spatial information, for both images used in experiments.

TABLE II. OVERALL ACCURACY (%) AND KAPPA COEFFICIENT OF CLASSIFIED IMAGES

Methods	Image 1		Image 2	
	Overall accuracy	Kappa coefficient	Overall accuracy	Kappa coefficient
SVM using only spectral information	87.56%	0.87	88.79%	0.88
The stacked features approach	94.13%	0.93	92.13%	0.91
The direct summation kernel	94.26%	0.93	92.38 %	0.92
The weighted summation kernel	94.48%	0.93	92.55%	0.92

TABLE III. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE WEIGHTED SUMMATION KERNEL FOR IMAGE 1. GLOBAL ACCURACY = 94.48%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	93,66	1,41	1,91	1,01	1,63	0,38
Green area	1,13	94,99	0,00	1,08	1,54	1,26
Tree	0,28	1,07	92,82	2,50	0,82	2,51
Soil	4,84	0,95	0,00	93,87	0,34	0,00
Building	0,01	1,16	2,69	0,47	95,67	0,00
Shadow	0,08	0,42	2,58	1,07	0,00	95,85

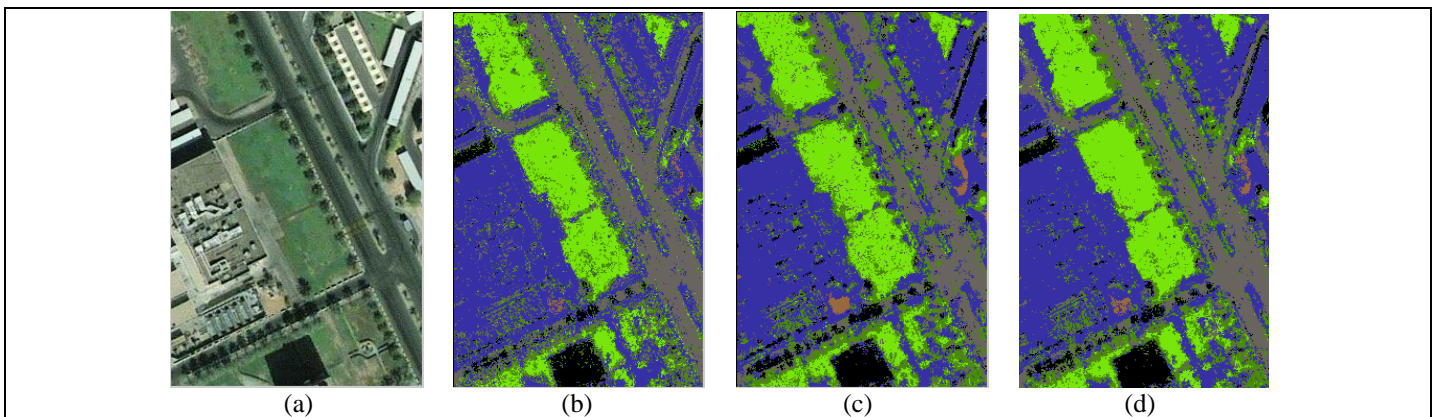


Fig. 7. (a) Original image 1, (b) Classification Map obtained using the stacked features approach, (c) Classification Map obtained using the direct summation kernel, (d) Classification Map obtained using the weighted summation kernel.

TABLE IV. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE WEIGHTED SUMMATION KERNEL FOR IMAGE 2.
GLOBAL ACCURACY = 92.55%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	89,36	2,04	1,92	1,50	3,32	1,86
Green area	5,13	92,21	0,00	1,03	1,54	0,09
Tree	1,18	1,52	93,15	1,92	0,03	2,20
Soil	1,75	1,13	0,64	93,04	3,44	0,00
Building	1,96	2,78	2,72	0,87	91,67	0,00
Shadow	0,62	0,32	1,57	1,64	0,00	95,85

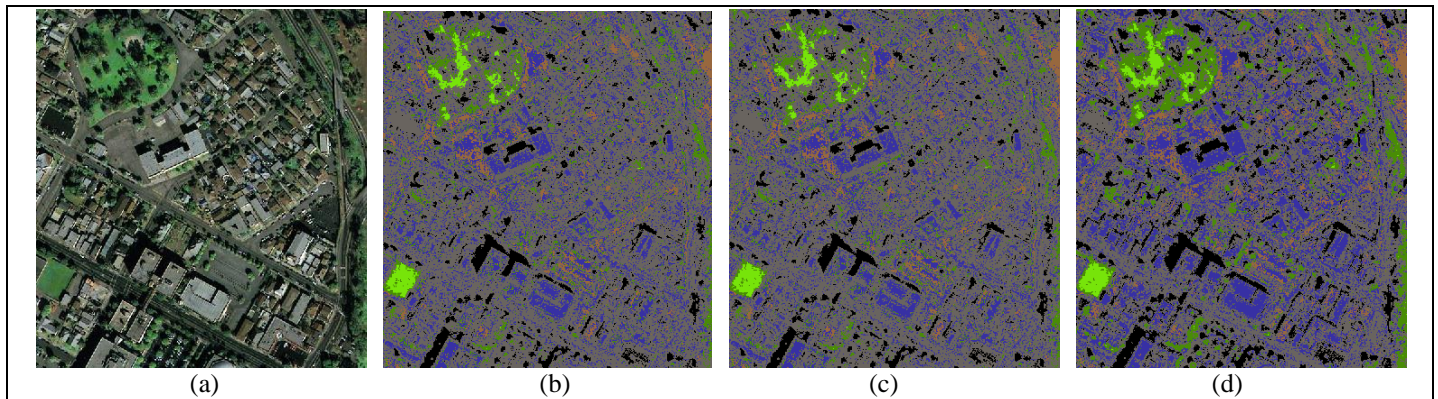


Fig. 8. (a) Original image 2, (b) Classification Map obtained using the stacked features approach, (c) Classification Map obtained using the direct summation kernel, (d) Classification Map obtained using the weighted summation kernel.

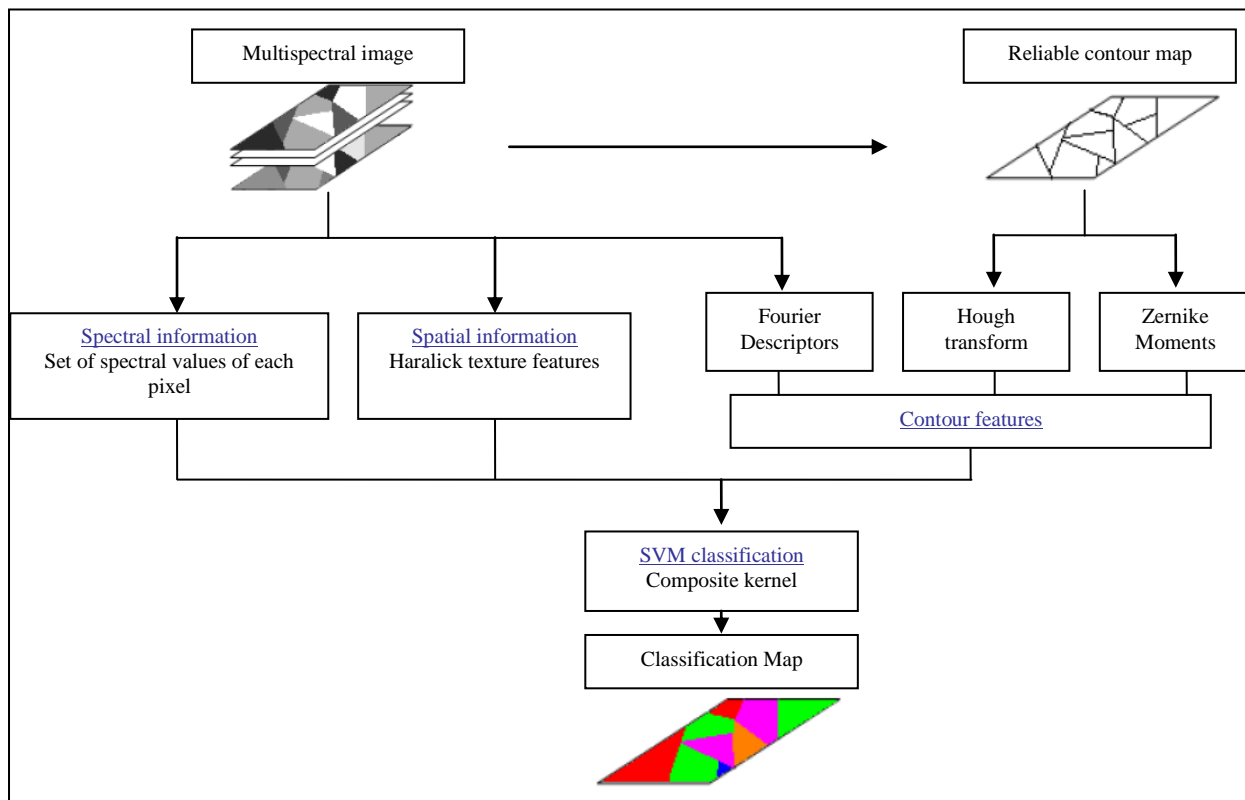


Fig. 9. A representative illustration of the second stage of the proposed workflow

C. Introducing Contour Information

In the second stage (represented by Fig.9.) we have started, like the first stage, with the extraction of the spectral and

spatial features, so we have computed Grey Level Co-occurrence Matrix (GLCM) to extract Haralick texture features that we have added to spectral information. But, before the

SVM classification, we have an additional step that consists on building a reliable contour map from which we have extracted contour descriptors specially Hough transform and Zernike moments, while Fourier descriptors are extracted directly from the original image.

1) Edge Detector Choice

Generally the edge detectors can be grouped into three major categories: the first one is the Early vision edge detectors (Gradient operators, e.g. the detectors of Sobel and Kirsch). The second category is Optimal detectors (e.g. the Canny algorithm, etc.). The third category is the Operators using parametric fitting models (e.g. the detectors of Haralick, Nalwa-Binford, Nayar, Meer and Georgescu, etc) [53].

The edge detection process is greatly eased if, instead the original images, «edge enhanced» ones are used. This inevitably leads to the use of some edge detectors from the second category.

In the present work, we have chosen to use Canny edge detector. John Canny has treated edge detection as signal processing problem and aimed to design the «optimal» edge detector. He formally has specified an objective function to be optimized and used this to design the operator.

The objective function was designed to achieve the following optimization constrains [54]:

- Maximize the signal to noise ratio in order to provide good detection.
- Achieve good localization to accurately mark edges.
- Minimize the number of responses to a single edge (non-edges are not marked).

2) Building a Reliable Contour Map

The Canny method finds edges by looking for local maxima of the gradient of the image. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges.

For simplicity and for illustrative purposes, we have used edge function in Matlab to extract contour map with the Canny method, and we have specified a scalar for thresh, this scalar value is used for the high threshold and 0.4*thresh is used for the low threshold. This scalar was varied with a step of 0.1 in the range [0, 1]. The Fig.10. Represents two values of threshold used for the first image.

For the choice of thresholds of the image contours that gives us a reliable contour map which will be used later in the classification process, we have adopted two measures proposed by Wiedemann [55], which are used for the evaluation of extraction methods roads from satellite images, these two measures are defined as follows:

Completeness = length of the reference contour in accordance with the extracted contour / length of the reference contour

Exactness = length of the extracted contour in accordance with the reference contour / length of the extracted contour.

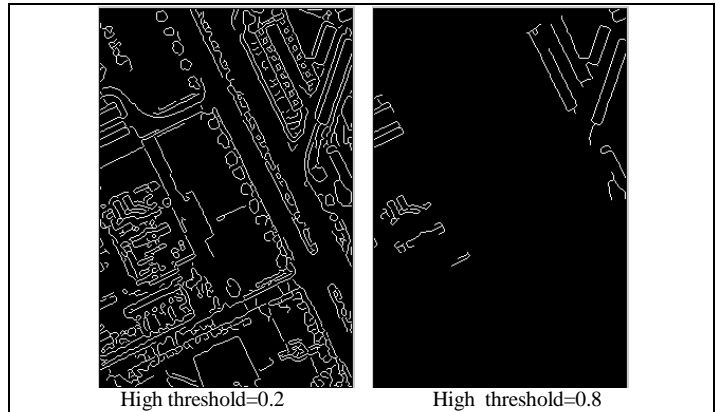


Fig. 10. example of contour map for image 1

The principle is to compare the contours of each threshold with the reference contours which are the contours of the SVM classification using the spectral and spatial information (Fig.11.).

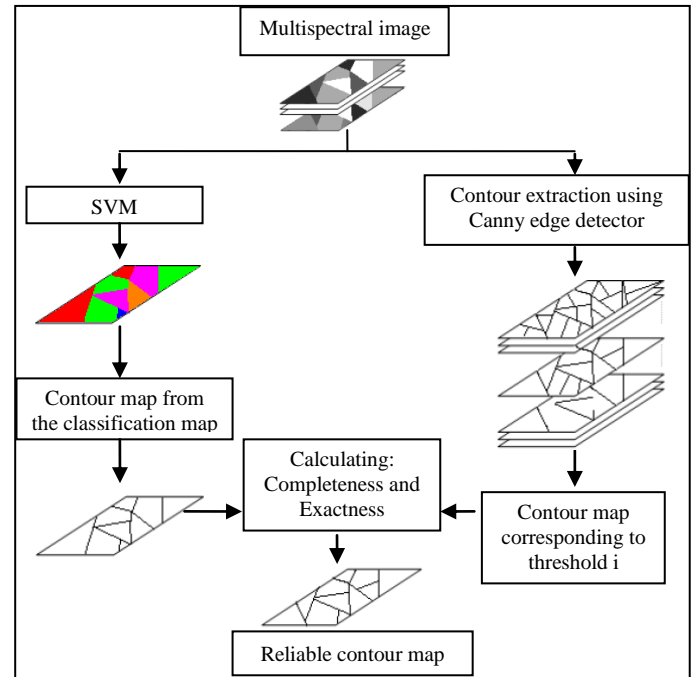


Fig. 11. Selecting reliable contour map

The comparison is made through the calculation of these measures. The constraint is that the selected threshold map is the one in which the extracted contours are the closest to the classification reference contours. The assessment method implemented in our study has a tolerance of a width of three pixels along the edges. The Fig.12. Represents a threshold evaluation for both images. The choice of thresholds of the image contours that gives us a reliable contour map that we have taken the one with a good both Completeness and Exactness, so we have chosen threshold 0.3 for image 1 and 0.4 for image 2 as we can see in Fig.12.

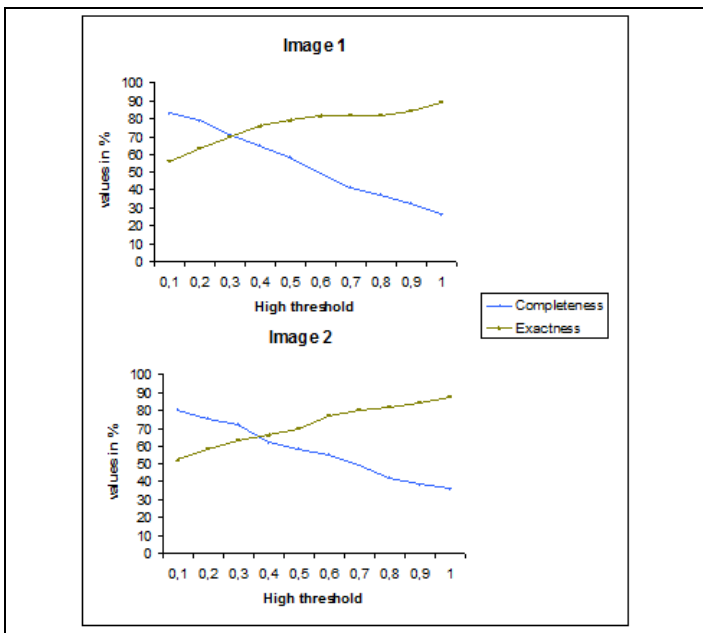


Fig. 12. threshold evaluation for the two images

3) Results

To combine spectral, spatial and contour information, we have used the extended weighted summation kernel in (38) that gave the best performance at the first stage of our experiments.

Where the μ_m are varied in the range [0, 1] to satisfy the condition $\sum_{m=1}^3 \mu_m = 1$. For simplicity and for illustrative purposes, all μ_m were the same for all classes in our experiments. The penalization factor in the SVM was tuned in the range $C = \{10^{-1} \dots 10^7\}$.

In this work, we have computed the participation of contour information in function of spectral and spatial information: $\mu_3 = 1 - (\mu_1 + \mu_2)$ and we have varied μ_1 and μ_2 with a step of 0.1 in the range [0, 1] to satisfy the condition $\sum_{m=1}^3 \mu_m = 1$.

We have used the Gaussian RBF kernel (28) (with $\sigma = \{10^{-1} \dots 10^3\}$) for all kernels. k_{spect} uses a spectral information, k_{spa} uses Haralick features while k_{cont} uses Fourier descriptors, Hough transform and Zernike moments.

The image (c) in Fig.13. and Fig.14. represent the reliable contour map used to compute contour descriptors' (Hough transform and Zernike moments); while (d) in Fig.13. and Fig.14. represent the classification map resulting by introducing contour (Fourier descriptors, Hough transform and Zernike moments) information with both spectral and spatial information.

A visual analysis of classification maps shows that it 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 V lists the accuracy estimates and kappa coefficient of the classification results, we can find different combination of descriptors used to characterize the contour information all models are compared numerically (overall accuracy, kappa coefficient).

Table VI and Table VII present respectively the confusion matrix results for SVM classification using the extended weighted summation kernel (38) based on spectral, spatial and contour information for both images used in experiments.

TABLE V. OVERALL ACCURACY (%) AND KAPPA COEFFICIENT OF CLASSIFIED IMAGES USING THE EXTENDED WEIGHTED SUMMATION KERNEL

Used Descriptors	Image 1		Image 2	
	Overall accuracy	Kappa coefficient	Overall accuracy	Kappa coefficient
Spectral + haralick features	94.48%	0.93	92.55%	0.92
Spectral + haralick features + FD	94.91%	0.93	92.88%	0.92
Spectral + haralick features + ZM	94.59%	0.93	92.68%	0.92
Spectral + haralick features + HT	94.49%	0.93	92.56%	0.92
Spectral + haralick features + FD +HT	95.06%	0.93	93.13%	0.92
Spectral + haralick features + FD + ZM	95.94%	0.94	93.98%	0.93
Spectral + haralick features + HT + ZM	95.81%	0.94	93.72%	0.93
Spectral + haralick features + FD + HT + ZM	96.17%	0.95	94.08%	0.93

The composite kernels offer excellent performance for the classification of multispectral satellite images by simultaneously exploiting both the spatial and spectral information. The weighted summation kernel allows a significant improvement of the classification performances when compared with the two other approaches. So the extended weighted summation kernel has been selected to introduce contour information.

The experimental results indicate a global accuracy value of 93.52%, the addition of contour information, described by the Fourier descriptors, Hough transform and Zernike moments, allows increasing the obtained global accuracy by 1.61% (using all descriptors) which is very promising. Although the Hough transform don't give a remarkable increasing of the overall accuracy, it preserves the edges in the obtained classification map.

TABLE VI. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE EXTENDED WEIGHTED SUMMATION KERNEL WITH ALL DESCRIPTORS FOR IMAGE 1. GLOBAL ACCURACY = 96.17 %

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	96,52	0,34	1,92	0,00	0,62	0,60
Green area	1,03	96,78	0,00	0,03	0,87	1,29
Tree	0,18	1,36	95,42	0,38	0,00	2,66
Soil	0,00	0,34	0,13	96,94	2,49	0,10
Building	1,94	1,16	0,81	0,08	96,01	0,00
Shadow	0,33	0,02	1,72	2,57	0,01	95,35

TABLE VII. CONFUSION MATRIX RESULTS (%) FOR SVM CLASSIFICATION USING THE EXTENDED WEIGHTED SUMMATION KERNEL WITH ALL DESCRIPTORS FOR IMAGE 2. GLOBAL ACCURACY = 94.08%

Class name	Asphalt	Green area	Tree	Soil	Building	Shadow
Asphalt	93,23	1,00	3,21	0,00	0,64	1,92
Green area	1,04	95,18	0,00	1,08	1,44	1,26
Tree	0,28	1,08	93,91	1,40	0,82	2,51
Soil	3,33	1,26	0,00	93,07	2,34	0,00
Building	1,41	1,06	0,41	2,36	94,76	0,00
Shadow	0,71	0,42	2,47	2,09	0,00	94,31

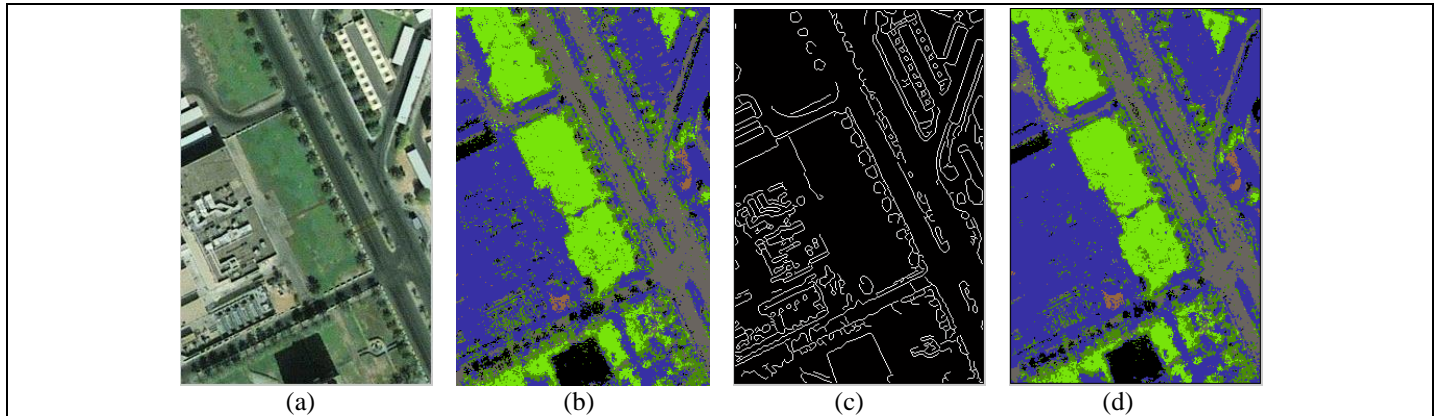


Fig. 13. (a) Original image 1, (b) Classification Map obtained using the weighted summation kernel, (c) A reliable contour map and (d) Classification Map obtained using the extended weighted summation kernel

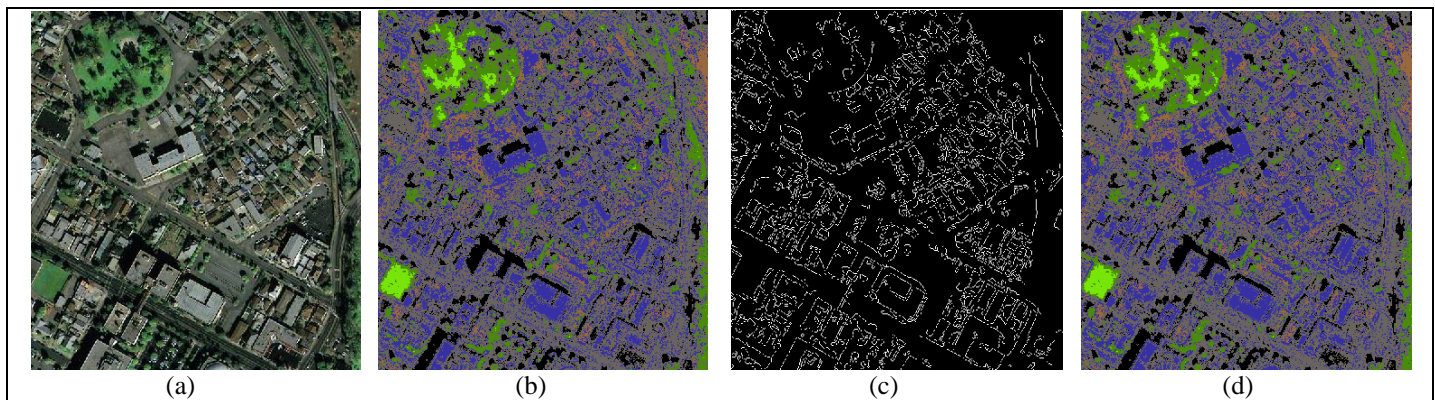


Fig. 14. (a) Original image 2, (b) Classification Map obtained using the weighted summation kernel, (c) the reliable contour map and (d) Classification Map obtained using the extended weighted summation kernel

VI. CONCLUSION AND FUTURE RESEARCH LINES

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. So an extended version of this kernel has been selected to introduce contour information (Fourier descriptors, Hough transform and Zernike moments). This approach exhibits flexibility to balance between the spectral, spatial and contour information as well as computational efficiency.

The proposed method is computationally expensive in comparison with a single kernel-based approach. In order to address this issue, we are planning on exploring the impact of reducing the original data set dimensionality before applying the proposed approach.

We are also planning to explore nonlinear combination methods, and the data-dependent combination methods which assign specific kernel weights for each data instance, to identify local distributions in the data and learn proper kernel combination rules for each region.

ACKNOWLEDGMENT

This work was funded by CNRST Morocco and CNRS France Grant under "Convention CNRST CRNS" program SPI09/12.

REFERENCES

- [1] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, 1968, IT vol. 14 no. 1 pp. 55-63.
- [2] D. A. Landgrebe, "Signal Theory Methods in Multispectral Remote Sensing," New York: Wiley 2003.
- [3] C. Samson "Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles": Thesis of doctorate, university of Nice-Sophia Antipolis 2000.
- [4] J.R.G. Townshend, "Land cover". *International Journal of Remote Sensing*, 1992 vol. 13 pp. 1319-1328.
- [5] F.G. Hall, J.R. Townshend, E.T. Engman, "Status of remote sensing algorithms for estimation of land surface state parameters." *Remote Sensing of Environment*, 1995 vol. 51 pp. 138-156.
- [6] 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.
- [7] C. Huang, L.S. Davis and J.R.G. Townshend, "An assessment of support vector machines for land cover classification." *International Journal of Remote Sensing*, 2002 vol. 23 pp. 725-749.
- [8] T. Kavzoglu, S. Reis, "Performance analysis of maximum likelihood and artificial neural network classifiers for training sets with mixed pixels." *GIScience and Remote Sensing*, 2008 vol. 45 pp. 330-342.
- [9] M. Pal, and P.M. Mather, "Support vector machines for classification in remote sensing." *International Journal of Remote Sensing*, 2005 vol. 26 pp. 1007-1011.
- [10] G. Zhu, and D.G. Blumberg, "Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel." *Remote Sensing of Environment*, 2002 vol. 80 pp. 233-240.
- [11] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, et al. "Comparing support vector machines with gaussian kernels to radial basis function classifiers." *IEEE Transactions on Signal Processing*, 1997 vol. 45 pp. 2758-2765.

- [12] X. Cao, J. Chen, H. Imura, O. Higashi, "A SVM-based method to extract urban areas from DMSP-OLS and SPOT VGT data", *Remote Sensing of Environment*, 2009 vol. 113 pp. 2205-2209.
- [13] J. Inglada, "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*, 2007 vol. 62 pp. 236-248.
- [14] A. Bekkari, S. Idbraim, D. Mammass and M. El yassa "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 2011.
- [15] M. Fauvel, J.A. Benediktsson, J. Chanussot and J.R. Sveinsson, "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles" *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 07*, Barcelona Spain 2007.
- [16] L. Lepisto, I. Kunttu, J. Autio, and A. Visa, "Classification method for colored natural textures using gabor filtering," *Image Analysis and Proc.*, pp. 397-401, Sept. 2003.
- [17] L. Lepisto, I. Kunttu, J. Autio, and A. Visa, "Rock image classification using non-homogeneous textures and spectral imaging," *Proc. of the WSCG*, 2003.
- [18] J.A. Palmason, J.A. Benediktsson, J.R. Sveinsson, and J. Chanussot, "Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis," *IEEE Trans., Int. Geosci. and Rem. Sens.*, vol. 1, July 2005.
- [19] A. Plaza, P. Martinez, R. Perez, and J. Plaza, "Spatial/ spectral endmember extraction by multidimensional morphological operations," *IEEE Trans., Int. Geosci. and Rem. Sens.*, vol. 40, no. 9, pp. 2025-2041, Sep 2002.
- [20] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Classification based marker selection for watershed transform of hyperspectral images," *IEEE Trans., Int. Geosci. and Rem. Sens. Symp.*, 2009.
- [21] G. Roussel, V. Achard, A. Alakian and J.C. Fort 'benefits of textural characterization for the classification of hyperspectral images', 2nd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010, pp:1-4.
- [22] G. Roussel 'Développement et évaluation de nouvelles méthodes de classification spatiale-spectrale d'images hyperspectrales', Theses Days ISAE Toulouse France, 2010.
- [23] M. Sharma, M. Markou and S. Singh 'evaluation of texture methods for image analysis' *Intelligent Information Systems Conference*, The Seventh Australian and New Zealand 2001, pp. 117-121.
- [24] W.Y. Chiu, and I. Couloigner, "Evaluation of incorporating texture into wetland mapping from multispectral images" *University of Calgary, Department of Geomatics Engineering, Calgary, Canada, EARSeL eProceedings* 2004.
- [25] R.M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification." *IEEE Transactions on Systems Man and Cybernetics*, 1973.
- [26] J.S. Weszka, C.R. Dyer, and A. Rosenfeld. "A Comparative Study of Texture measures for Terrain Classification." *IEEE Transactions on Systems Man and Cybernetics*, 1976.
- [27] R.W. Conners, and C.A. Harlow, "A Theoretical Comparison of Texture Algorithms." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1980.
- [28] P. Gong, D. J. Marceau and P. J. Howarth 'A Comparison of Spatial Feature Extraction Algorithms for Land-Use Classification with SPOT HRV Data', *Remote Sensing Environ.*, 1992. vol. 40 pp. 137-151.
- [29] V. Arvis, C. Debain, M. Berducat and A. Benassi, "Generalization of the cooccurrence matrix for colour images: application to colour texture classification" *Journal Image Analysis and Stereology*, 2004 vol. 23 pp. 63-72.
- [30] H. Ming-Kuel, "Visual pattern recognition by moment invariants", *IRE trans. on Information Theory*, 1962 vol. 8 pp. 179-187.
- [31] M. Teague, "Image analysis via the general theory of moments", *Journal of the Optical Society of America*, 1980 vol. 70 no. 8 pp. 920-930.

- [32] R. L. Cosgriff, "Identification of shapes", Ohio State Univ. Res. Foundation, Columbus, 1960 Rep. 820-11, ASTIA AD 254 792.
- [33] D. L. Fritzsche, "A systematic method for character recognition", Ohio State Univ. Res. Foundation, Columbus, 1961 Rep. 1222-4, ASTIA AD 268 360.
- [34] G. H. Granlund, "Fourier preprocessing of hand printed character recognition", IEEE Trans. Comp., 1972 C-21 pp. 195-201.
- [35] C. T. Zhan, R. S. Roskies, "Fourier descriptors for plane closed curves", IEEE Trans. Comp., 1972 C-21 pp. 269-281.
- [36] L. Yang, R. Prasad, "Recognition of line-drawing based on generalized Fourier descriptors", International Conference on Image Processing and its Applications, 1992 pp. 286-289.
- [37] C. Adamos, W. Faig "Hough Transform in Digital Photogrammetry". In: International Archives of Photogrammetry and Remote Sensing, Washington, USA, 1992 vol. 29, Part B3 pp. 250-254.
- [38] D. H. Ballard, "Generalizing the Hough Transform to detect arbitrary shapes". Pattern Recognition, 1981 vol. 13 no. 2 pp. 111- 122.
- [39] R. D. Duda, P. E. Hart, "Use of the Hough Transform to detect lines and curves in pictures. " Communication of the ACM, 1972 vol. 15 no. 1 pp. 11-15.
- [40] J. D. Klingensmith, R. Shekhar, and D. Geoffrey, "Evaluation of Three-Dimensional Segmentation Algorithms for the Identification of Luminal and Medial-Adventitial Borders in Intravascular Ultrasound Images, " IEEE Transactions on Medical Imaging, 2000 vol.19 pp. 996-1011.
- [41] M. S. Nixon, A. S. Aguado, "Feature extraction & image processing, " Academic Press Inc, 2nd revised edition 2007.
- [42] R. Roman-Roldan, J. Francisco-Gomez-Lopera, C. Ataallah, J. Martinez-Aroza, P. L. Luque-Escamilla, "A measure of quality for evaluating methods of segmentation and edge detection, Pattern Recognition," 2001 vol. 34 pp. 969-980.
- [43] W. C. Siu, "Efficient computation of moments for pattern recognition, " IEEE Pacific Rim Conference on Communications Computers and Signal Processing, 1991 vol. 2 pp. 589-592.
- [44] R. Unnikrishnan, C. Pantofaru and M. Hebert, "A measure for objective evaluation of image segmentation algorithms, " IEEE Workshops on Computer Vision and Pattern Recognition, 2005 pp. 34-34.
- [45] Y. J. Zhang, "A review of recent evaluation methods for images segmentation, " International Symposium on Signal Processing and its Applications, 2001 vol. 1 pp. 148-151.
- [46] L. Chapel, " 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 2007.
- [47] S. Aseervatham " 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 2007.
- [48] O. Bousquet, "Introduction au Support Vector Machines (SVM)", Center mathematics applied, polytechnique school of Palaiseau 2001. <http://www.math.u-psud.fr/~blanchard/gtsvm/index.html>.
- [49] M. Fauvel, J. Chanussot and J. A. Benediktsson "A Combined Support Vector Machines Classification Based on Decision Fusion" IEEE International Geoscience and Remote Sensing Symposium, IGARSS 06, Denver, USA 2006.
- [50] B. Mak, J. Kwok, and S. Ho, "A study of various composite kernels for kernel eigenvoice speaker adaptation," in IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP04, 2004 vol. 1.
- [51] J.-T. Sun, B.-Y. Zhang., Z. Chen, Y.-C. Lu., C.-Y. Shi, and W. Ma, "GECKO: A method to optimize composite kernels for web page classification," in IEEE/WIC/ACM International Conference on Web Intelligence, WI04, 2004 vol. 1 pp. 299–305.
- [52] SVMlight Version: 6.02 Developed at University of Dortmund, Informatik, AI-Unit Collaborative Research Center on 'Complexity Reduction in Multivariate Data' (SFB475). 2008 <http://svmlight.joachims.org/>
- [53] D. Ziou, and S. Tabbone, "Edge Detection Techniques – An Overview. TR-195", Département de Math et Informatique, Université de Sherbrooke, Québec, Canada, 1997 pp. 1-41.
- [54] J. Canny, "A computational approach to edge detection". IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986 vol. 8 no. 6, pp. 679-698.
- [55] C. Wiedemann, C. Heipke, H. Mayer, "Empirical evaluation of automatically extracted road axes", In: CVPR Workshop on Empirical Evaluation Methods in Computer Vision, California, , 1998 pp. 172–187.