Special Issue of International Journal of Computer Applications (0975 – 8887)
on Software Engineering, Databases and Expert Systems – SEDEXS, September 2012

Classification of High Resolution Urban satellites Images using SVM and Haralick Features with a Hybrid Median Filter

Aissam Bekkari, Soufian Idbraim, Azeddine Elhassouny, Driss Mammass and Mostapha 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).

For this purpose, we propose a methodology exploiting a composite kernel that easily combines multi-spectral features, Haralick texture features and Hybrid Median Filter, with different window sizes.

The proposed approach was tested on common scenes of urban imagery. The result shows that the combined use of spectral and texture information together significantly improved the accuracy of satellite image classification.

Keywords
SVM, Composite Kernels, Haralick features, Hybrid Median Filter, Satellite image, Spectral and spatial information, GLCM.

1. 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.

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [1-4]. 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), which is a group of advanced machine learning algorithms, have seen increased use in land cover studies [7, 8].

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 [9,10].

Very important to know is that 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.

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. We start with the extraction of spectral and spatial information (Haralick texture features [11] and Hybrid Median Filter). Then, we apply the SVM classification to the result file.

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, and the Hybrid Median Filter used in experimentations. In section3, we give outlines on the used classifier: Support Vector Machines (SVM). In section4 the results are presented as well as the stating of numerical evaluation. Finally, conclusions are given in section5.

2. 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 [12] this allows a good classification based on the spectral signature of each area. However, this does not take into account the spatial information represented by the various structures in the image.

2.2 Spatial information

2.2.1 Haralick Features
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) [13] 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.

The size of the co-occurrence matrix equals to the number of the image gray levels, also the dynamics of the image is usually small (typically, 8 gray levels) in order not to work with too large matrices.

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 [14], however only some of these features are in wide use.

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.

2.2.2 Hybrid Median Filter

The hybrid median filter [15] is a modification of median filter. This filter is also called as corner preserving median filter that is a three-step ranking operation. In a 5X5 pixel neighbourhood, as example, pixels can be ranked in two different groups.

$$\begin{pmatrix} D & * & R & * & D \\ * & D & R & D & * \\ R & R & C & R & R \\ * & D & R & D & * \\ D & * & R & * & D \end{pmatrix}$$

Fig1: 5X5 window for Hybrid median Filter

The median values of the 45° neighbours forming a "×" (pixels marked as D in Fig1) and the 90° neighbours forming a "+" (pixels marked as R in Fig1) are compared with the central pixel and the median value of that set is then saved as the new pixel value. The three-step ranking operation does not impose a serious computational penalty as in the case of median filter. Each of the ranking operations is for a much smaller number of values than used in a square region of the same size. For example, the 5 pixel wide neighbourhood used in the examples contains either 25 (in the square neighbourhood) which must be ranked in the traditional method. In the hybrid method, each of the two groups contains only 9 pixels, and the final comparison involves only three values. Even with the additional logic and manipulation of values, the hybrid method is faster than the conventional median. This median filter overcomes the tendency of median and truncated median filters to erase lines which are narrower than the half width of the neighbourhood and to round corners.

3. SVM CLASSIFICATION

In this section, we briefly describe the general mathematical formulation of SVMs introduced by Vapnik [16, 17]. 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 $\mathrm{x_i} \in \mathrm{R}^n$, are available with their corresponding labels yi = ±1, S = {(xi, yi) | 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 [18]. Noting $\mathrm{w} \in \mathrm{R}^n$ 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 $\langle \mathrm{w}, \mathrm{x} \rangle$ is the inner product between w and x. If x $\notin$ Hp then f(x) = $\langle \mathrm{w}, \mathrm{x} \rangle$ + 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:

$$\min \left[\frac{\|w\|^2}{2} \right] \quad \text{subject to } y_i(\langle w, x_i \rangle + b) \geq 1 \ \ \forall i \in [1, l] \tag{7}$$

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

$$\min \left[\frac{\|w\|^2}{2} + C \sum_{i=1}^{l} \xi_i \right] \quad \text{subject to } y_i(\langle w, x_i \rangle + b) \geq 1 - \xi_i, \xi_i \geq 0 \ \ \forall i \in [1, l] \tag{8}$$

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:

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle$$
$$\text{subject to } 0 \le \alpha_i \le C \qquad \forall i \in [1, l] \tag{9}$$
$$\sum_{i=1}^{l} \alpha_i y_i = 0$$

Finally:

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i \tag{10}$$

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:

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

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*:

$$\begin{aligned} \Phi : R^n &\to F \\ x &\mapsto \Phi(x) \end{aligned} \tag{12}$$

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:

$$\langle \Phi(x_i), \Phi(x_j) \rangle = k(x_i, x_j) \tag{13}$$

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:

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

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

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 [19]. 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.

4. EXPERIMENTAL RESULTS

4.1 Experiments and Data

The proposed workflow has two main tasks, we start with the extraction of spectral information and spatial information and then the result will be used as an input to SVM classifier.

To use jointly spatial and spectral information, we chose to go through the definition of a kernel. In [20], several kernels are proposed to include spatial information. The weighted sums of kernels provide the best results for classification.

They also allow to control the influence of each type of information:

$$k_{\mu}(x, y) = \mu k_{spectral}(x, y) + (1 - \mu) k_{spatial}(x, y) \quad \text{with } 0 \le \mu \le 1 \tag{16}$$

The parameter μ will be chosen at the learning phase, it varied in steps of 0.1. For simplicity and for illustrative purposes, μ was the same for all the classes in our experiments. The penalization factor in the SVM was tuned in the range *C* = $\{10^{-1} \ldots 10^{7}\}$.

We use a RBF kernel (15) (with $\sigma = \{10^{-1} \ldots 10^{3}\}$) for the two kernels. $k_{spectral}$ uses a spectral information while $k_{spatial}$ uses Haralick features and Hybrid Median Filter.

The Hybrid Median Filter is calculated with three window sizes 3X3, 5X5 and 7X7. So if we take the 5X5 window represented in Fig1 as an example, first we find the median MR of the pixels marked as R in Fig1 and the central pixel C in this window, and then we find the median MD of the pixels marked as D in Fig1 and the central pixel C in this window.

The median value of the result set is then saved as the new pixel value, and it's the same for other windows.

Concerning the data, we have used a multispectral satellite image (Quickbird) represented in Fig 2 (a), with a size of 600 by 800 pixels; finally we will have for this image 4 131 individuals (pixels) for learning, 4 952 for validation and 480 000 for classification, divided on sex classes as follow (Table 1):

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

4.2 The results

In this section, we present the obtained results, first, we compute Grey Level Co-occurrence Matrix (GLCM) to extract Haralick texture features that we add to spectral information, thereafter, we use SVM for the classification of the result file.

To the file that contains spectral and spatial information obtained from the original image ((a) in Fig 2) we apply an SVM classification with composite kernel (16).

The classification map presented on Fig 2 (b), is obtained when the classification is performed using only spectral information, and the classification map presented on Fig 2 (c), is obtained when the classification is performed using spectral information with Haralick features.

The fusion of spectral information and Hybrid Median Filter with three window sizes 7X7, 5X5 and 3X3 give us respectively the classification map (d), (e) and (f) presented on Fig 2.

The results have progressed with the combined use of spectral and spatial information. In addition, a visual analysis of the classification maps shows those areas are more homogeneous for the map obtained with the proposed SVM using spectral and Haralick features.

The classification maps obtained by the fusion of the spectral information and Hybrid Median Filter are less noisy and the classification performances are increased globally as well as almost all the classes.

In conclusion both of those methods matche well with an urban land cover map in terms of smoothness of the classes; and it also represents more connected classes mainly when using 3X3 window in Hybrid Median, it gives us approximately the same result when using Haralick Features in terms of global accuracy. Table 2 lists the accuracy estimates for the study area, all models are compared numerically (overall accuracy).

Table 2. Overall accuracy (%) of classified image

<table>
<tr><th colspan="2">Method</th><th>Accuracy</th></tr>
<tr><td colspan="2">Spectral-SVM
(μ=1)</td><td>88.79%</td></tr>
<tr><td rowspan="4">Spectral-Spatial SVM</td><td>Haralick Features</td><td>92.13%</td></tr>
<tr><td>Hybrid Median Filter with 3X3 pixel neighbourhood</td><td>92.08%</td></tr>
<tr><td>Hybrid Median Filter with 5X5 pixel neighbourhood</td><td>91.20%</td></tr>
<tr><td>Hybrid Median Filter with 7X7 pixel neighbourhood</td><td>90.62%</td></tr>
</table>

5. CONCLUSION

Addressing the classification of high resolution satellite images from urban areas, we have presented a composite kernel taking simultaneously the spectral and the spatial information into account (the spectral values, Haralick features and Hybrid Median Filter).

The two methods give us a significant improvement of the classification performances when compared with the classification using spectral information only, however it remains to improve even more these results.

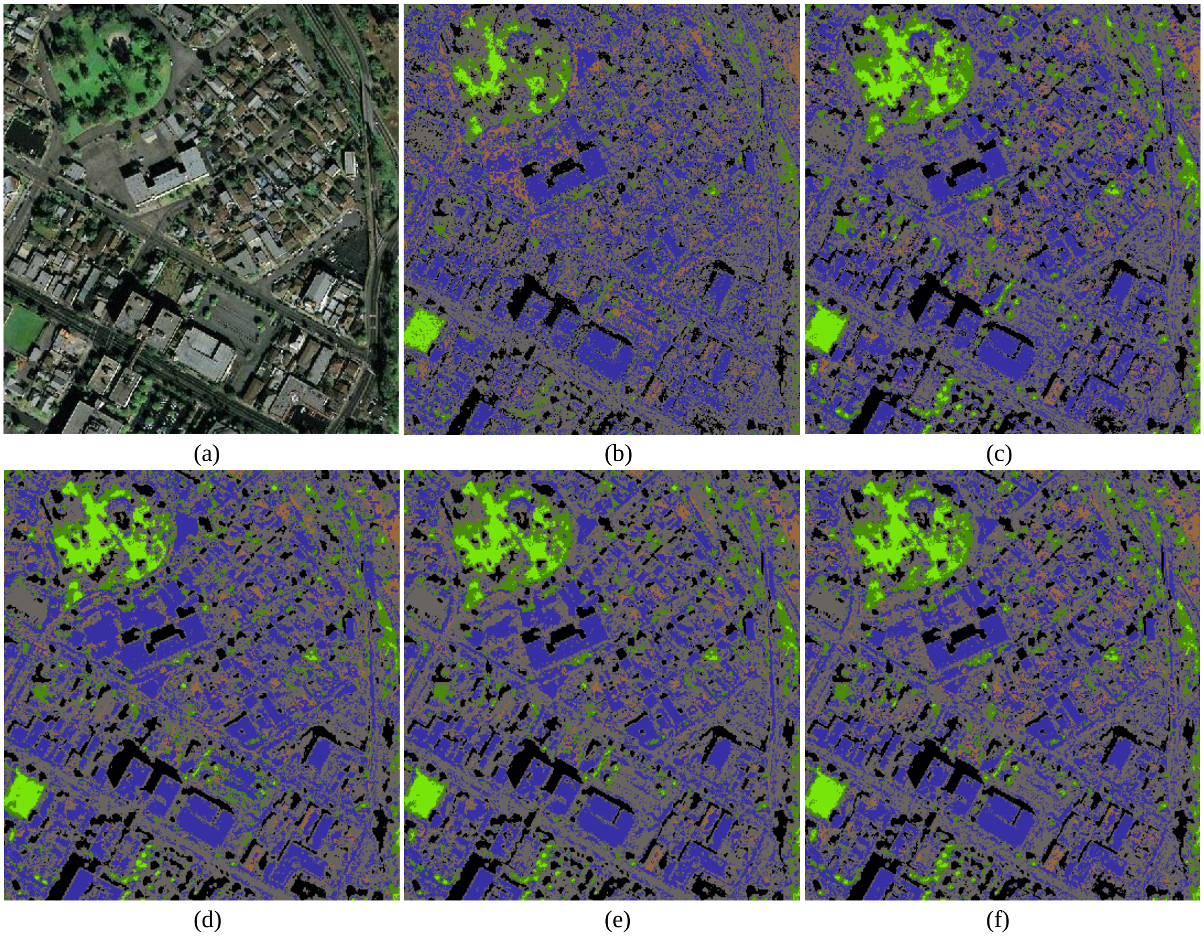

Fig2: (a) Original image, (b) Classification Map obtained using only spectral information, (c) Classification Map obtained using spectral information and haralick features and (d), (e) and (f) are Classification Maps obtained using spectral information and Hybrid Median Filter with respectively 7X7 window, 5X5 window and 3X3 window

6. ACKNOWLEDGMENTS

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

7. 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] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Q.Yang, 2011 "A Hybrid Median Filter for Enhancing Dim Small Point Targets and Its Fast Implementation" in IEEE International Conference on Multimedia and Signal Processing CMSP, 1:239 – 242.

[16] 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.

[17] 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.

[18] 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.

[19] J.-T. Sun, B.-Y. Zhang, Z. Chen, Y.-C. Lu, C.-Y. Shi, and W. Ma, (Sept 2004) "GE-CKO: A method to optimize composite kernels for web page classification," in IEEE/WIC/ACM International Conference on Web Intelligence, WI04, 1:299–305.

[20] Camps-Valls G., Gomez-Chova L., Munoz-Mari J., Vila-Francés J., and Calpe-Maravilla J. January 2006 "Composite kernels for hyperspectral image classification." IEEE Geoscience Remote Sensing Letters, 3(1):93–97