SVM and Haralick Features for Classification of High Resolution Satellite Images from Urban Areas

Aissam Bekkari[1], Soufiane Idbraim[1], Azeddine Elhassouny[1], Driss Mammass[1], Mostafa El yassa[1], and Danielle Ducrot[2]

[1] IRF – SIC Laboratory, Faculty of Sciences B.P. 80 000 Agadir – Morocco
{a_bekkari,soufianeidbraim,info_azeddine}@yahoo.fr, mammass@univ-ibnzohr.ac.ma, melyass@menara.ma
[2] Cesbio bpi 2801 31401 Toulouse cedex 9 France
danielle.ducrot@cesbio.cnes.fr

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 allowing to combine these two informations using a combination of multi-spectral features and Haralick texture features as data source with composite kernel. The proposed approach was tested on common scenes of urban imagery. The results allow a significant improvement of the classification performances when compared with the two sets of attributes used separately. The experimental results indicate an accuracy value of 93.29% which is very promising.

Keywords: SVM, composite kernel, Haralick features, Satellite image, Spatial and spectral 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 becomes accessible and increasingly frequent [1].

Several classification algorithms have been developed since the first satellite image was acquired in 1972 [2-4]. Recently, 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 [5, 6]. 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 [7].

A. Elmoataz et al. (Eds.): ICISP 2012, LNCS 7340, pp. 17–26, 2012.

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.

The proposed method consists in combining spatial and spectral information to obtain a better classification. We start with the extraction of spectral and spatial information. Then, we apply the SVM classification to the result file. Experimental results are provided and comparisons with a spectral classification and spatial classification are made to illustrate that the method is able to find better classes.

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 section 3, we give outlines on the used classifier: Support Vector Machines (SVM). In section 4, the results are presented with the used kernel defined as well as the stating of numerical evaluation. Finally, conclusions are given in section 5.

2 Extraction of Information and Classification

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 [8] 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. Many approaches were developed for texture analysis. Grey-Level Co-occurrence Matrix (GLCM) [9] 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 uses these matrices to develop a number of spatial indices that are easier to interpret. He 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 (f_1 to f_{14}) described by Haralick et al [10], however only some of these features are in wide use. Wezska et al [11] used four of Haralick features (f_1 , f_2 , f_5 , f_8). Conners and Harlow [12] use five features (f_1 , f_2 , f_3 , f_4 , f_5).We found that these five features are commonly used seen that the fourteen are much correlated with each other, and that the five sufficed to give good results in classification [13].

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.

2.3 SVM Classification

SVM is a group of advanced machine learning algorithms that have seen increased use in land cover studies; it generally produced more accurate results than other algorithms (decision trees and neural networks).

In this section we briefly describe the general mathematical formulation of SVMs introduced by Vapnik [14]. 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.

2.3.1 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 [15]. Noting w∈ R^n as the normal vector of the hyperplane and b ∈ 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 $y = \mathrm{sgn}(f(x))$.

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:

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

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:

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

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:

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

Where x_u is an unseen sample.

2.3.2 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:

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

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:

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

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

In experiments we used Gaussian RBF kernel (15) which is commonly used in classification of remotely sensed images.

2.3.3 Multiclass SVMs

SVMs are designed to solve binary problems where the class labels can only take two values: ±1. For a classification of remotely sensed images, several classes are usually of interest. Various approaches have been proposed to address this problem [16]. 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.

The *pairwise classification* has shown to be more suitable for large problems [15, 16]. Even though the number of the used classifiers is larger than for *the one versus the rest* approach, the whole classification problem is decomposed into much simpler ones. Therefore, this second approach was used in our experiments.

2.4 The Proposed Workflow

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 [17], 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:

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

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 $C = \{10^{-1}\ldots10^{7}\}$. 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.

3 Experimentations and Results

3.1 The Data

The first image used in classification is a sample of high resolution Quickbird satellite image. Its size is 240x360 pixels. It represents scene urban areas. We dispose of four spectral bands: blue, green, red and near infrared. We can see in Fig.1 (a) a representation of this image.

The second test image is another sample of Quickbird satellite image with exactly the same properties except the size, 500x280 pixels. The scene does contain also urban areas. The original image is represented in Fig.2 (a).

We will have two files for each image, "TrainFile.dat" and "TestFile.dat" respectively for learning and for classification, divided on sex classes as described in Table 1.

3.2 The Results

The classification maps presented on (b) respectively in Fig. 1 and Fig. 2, are obtained when the classification is performed using the spatial information only (Haralick features). We can note the appearance of misclassifications. When the classification is performed using the spectral information only, we obtain the corresponding classification maps which are presented on (c) respectively in Fig. 1 and Fig. 2. These results appear as noisy as the spatial information that is not taken into account.

The fusion of the spectral and the spatial features give us the classification maps presented on (d) respectively in Fig. 1 and Fig. 2.The classification maps are 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 summarizes the results obtained using the SVM classification with Gaussian RBF kernel. These values were extracted from the confusion matrix. The overall accuracy is the percentage of correctly classified pixels. Kappa coefficient is another criterion classically used in remote sensing classification to measure the degree of agreement and takes into account the correct classification that may have been obtained “by chance” by weighting the measured accuracies.

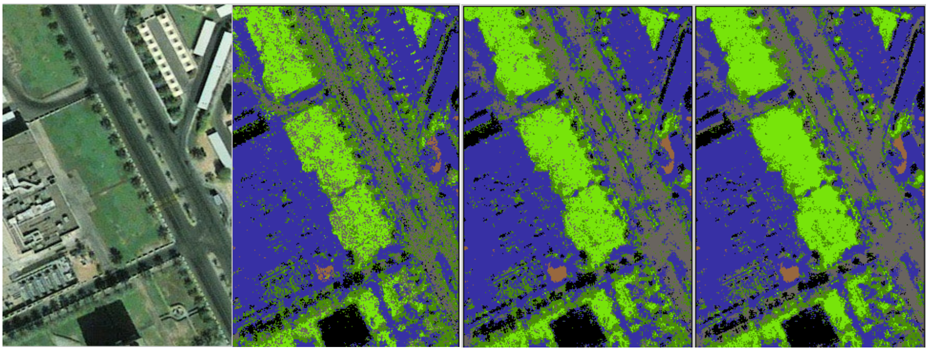
The use of this composite kernel (16) gives good classification results for the overall accuracy and the Kappa coefficient. Moreover, with all of the accuracies over 90%, this composite kernel seems also promising for the classification of remotely sensed images.

Table 1. Different classes

Class N°	*Class name*	***Train samples*** *Image 1*	*Image 2*
1	*Asphalt*	1 592	753
2	*Green area*	2 252	1 680
3	*Tree*	880	519
4	*Soil*	176	1 387
5	*Building*	4 217	1 282
6	*Shadow*	1 280	808
Total		10 397	6 429

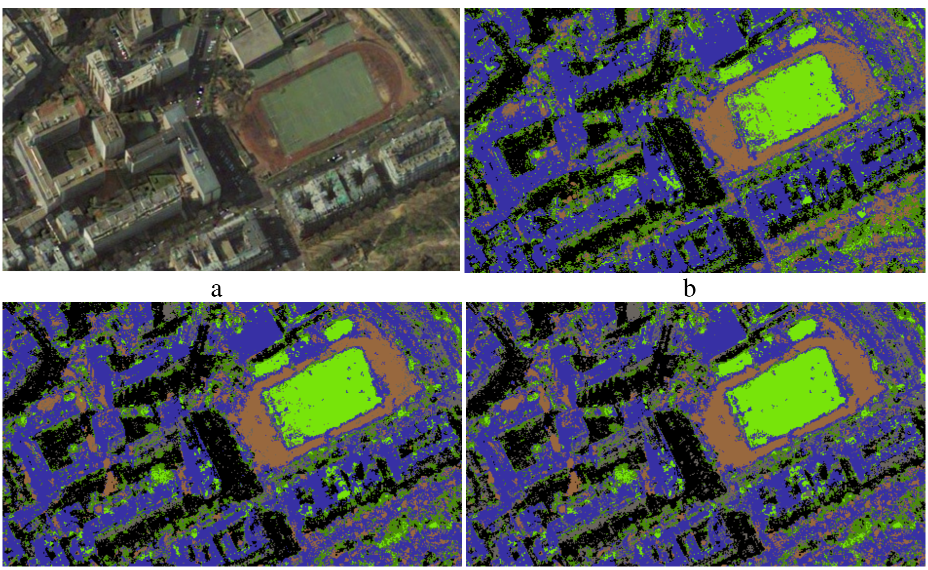
Table 2. Classification accuracies for the classified images

	Image 1			*Image 2*		
Methods	*SVM spatial*	*SVM spectral*	*SVM Spectral & spatial*	*SVM spatial*	*SVM spectral*	*SVM Spectral & spatial*
Overall accuracy	83.19%	87.27%	93.68%	85.24%	88.02%	92.90%
Kappa coefficient	0.85	0.89	0.93	0.86	0.89	0.92



a b c d

Fig. 1. (a) Original image, (b) Classification Map obtained with the classical RBF kernel using only spatial information, (c) Classification Map obtained with the classical RBF kernel using spectral information only and (d) map classification obtained with the proposed kernel



a b c d

Fig. 2. (a) Original image, (b) Classification Map obtained with the classical RBF kernel using only spatial information, (c) Classification Map obtained with the classical RBF kernel using spectral information only and (d) map classification obtained with the proposed kernel

4 Conclusion

Addressing the classification of high resolution satellite images from urban areas, we have presented an algorithm taking simultaneously the spectral and the spatial information into account. This is achieved by concatenating the two vectors of attributes (the spectral values and the Haralick features).

This data combination allows a significant improvement of the classification performances when compared with the two sets of attributes used separately.

As a perspective of this work, we will be concentrating on the study of the kernel choice in order to determine the appropriate one, for this type of image classification.

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

References

1. Samson, C.: Contribution à la classification des images satellitaires par approche variationnelle et équations aux dérivées partielles: Thesis of doctorate, University of Nice-Sophia Antipolis (2000)
2. Townshend, J.R.G.: Land cover. International Journal of Remote Sensing 13, 1319–1328 (1992)
3. Hall, F.G., Townshend, J.R., Engman, E.T.: Status of remote sensing algorithms for estimation of land surface state parameters. Remote Sensing of Environment 51, 138–156 (1995)
4. Lu, D., Weng, Q.: A survey of image classification methods and techniques for improving classification performance. International Journal of Remote Sensing 28, 823–870 (2007)
5. Pal, M., Mather, P.M.: Support vector machines for classification in remote sensing. International Journal of Remote Sensing 26, 1007–1011 (2005)
6. Zhu, G., Blumberg, D.G.: Classification using ASTER data and SVM algorithms: The case study of Beer Sheva, Israel. Remote Sensing of Environment 80, 233–240 (2002)
7. Scholkopf, B., Sung, K., Burges, C., Girosi, F., Niyogi, P., Poggio, T., et al.: Comparing support vector machines with gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing 45, 2758–2765 (1997)
8. Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R.: Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profiles. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2007, Barcelona Spain (2007)
9. Chiu, W.Y., Couloigner, I.: Evaluation of incorporating texture into wetland mapping from multispectral images. University of Calgary, Department of Geomatics Engineering, Calgary, Canada, EARSeL eProceedings (2004)
10. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. IEEE Transactions on Systems Man and Cybernetics (1973)
11. Weszka, J.S., Dyer, C.R., Rosenfeld, A.: A Comparative Study of Texture measures for Terrain Classification. IEEE Transactions on Systems Man and Cybernetics (1976)
12. Conners, R.W., Harlow, C.A.: A Theoretical Comaprison of Texture Algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence (1980)

13. Arvis, V., Debain, C., Berducat, M., Benassi, A.: Generalization of the cooccurrence matrix for colour images: application to colour texture classification. Journal Image Analysis and Stereology 23, 63–72 (2004)
14. Aseervatham, S.: 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)
15. Bousquet, O.: Introduction au Support Vector Machines (SVM). Center mathematics applied, polytechnique school of Palaiseau (2001), http://www.math.u-psud.fr/~blanchard/gtsvm/index.html
16. Fauvel, M., Chanussot, J., Benediktsson, J.A.: A Combined Support Vector Machines Classification Based on Decision Fusion. In: IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2006, Denver, USA (2006)
17. Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Francés, J., Calpe-Maravilla, J.: Composite kernels for hyperspectral image classification. IEEE Geoscience Remote Sensing Letters 3(1), 93–97 (2006)