

ADAPTIVE PIXEL NEIGHBORHOOD DEFINITION FOR THE CLASSIFICATION OF HYPERSPECTRAL IMAGES WITH SUPPORT VECTOR MACHINES AND COMPOSITE KERNEL

Mathieu Fauvel^{*◇}, Jocelyn Chanussot^{*}

Jon Atli Benediktsson[◇]

^{*}GIPSA-lab, Signal and Image Dept.
Grenoble Institute of Technology
BP 46 - 38402 St Martin d'Herès - FRANCE

[◇]University of Iceland
Dept. of Electrical and Computer Eng.
Hjardarhagi 2-6, 107 Reykjavik-ICELAND

ABSTRACT

The pixel-wise classification of hyperspectral images with a reduced training set is addressed. The joint use of the spectral and the spatial information is investigated. The spectral information simply consists of the spectral value of each pixel. For the spatial information, we use an *area filter* to simplify the image and extract consistent connected components. These components are used to define an adaptive neighborhood for each pixel of the image. The vector median value of each component is defined as a spatial feature for the classification. Support Vector Machines are used for the classification and a composite kernel is used to combine both the spatial and the spectral information. Experiments are conducted on AVIRIS hyperspectral data. The proposed approach provides significant improvements in terms of classification accuracy when compared with a standard statistical method (maximum likelihood) and with a SVM classifier using the spectral information alone. Robustness with respect to the size of the training set is also investigated.

Index Terms— Support vectors machines, spatial information, hyperspectral data, kernel function, area self-complementary filter.

1. INTRODUCTION

The pixel-wise classification of hyperspectral data using both spectral and contextual information is investigated in this paper. It is well known that contextual information is useful for the classification of hyperspectral images [1]. Such information is usually modeled in a statistical framework using Markov Random Field theory (MRF). For instance, a recursive Bayesian classifier is proposed in [2], where the joint prior probabilities of the classes of each pixel and its spatial neighbors being modeled by MRF. Improved classification performances are obtained, the gain over standard methods being significant when the training set remains limited. Advance approaches can be found in [3], for the regularization of remote sensing images. These approaches are based on a statistical modelling of the inter-pixel dependency, where the neighborhood of every pixel is defined by a fixed set such as a 3×3 square. This strategy fails when considering pixels that are close to the border of a structure: The fixed shape neighborhood then includes pixels from different structures. For example, as shown in Fig. 1.(a), the classification of the marked pixel (roof) may be biased by the neighboring pixels that actually

belong to the street. To overcome this problem, an *adaptive neighborhood* should be defined for each pixel. Furthermore, assuming that relevant structures have a sufficient area, this adaptive neighborhood should include a large number of pixels. In this paper, we thus propose to define the neighborhood of one given pixel as the connected zone resulting of a self-complementary area filtering. For each pixel, the spatial feature used for the classification is extracted from the neighborhood of pixels.

Another problem usually appears with MRF based methods and hyperspectral data: the problem of parameters estimation. When dealing with hyperspectral data, each pixel-vector \mathbf{x} is composed of several components ($\dim(\mathbf{x}) > 100$). In such a high dimensional space, the statistical estimation becomes a difficult task [4]. For instance, the required number of training pixels for a reliable estimation is related to the square of the dimensionality for a quadratic classifier (*e.g.*, the Gaussian Maximum Likelihood) [5]. In remote sensing applications, only limited ground-truth information is usually available. As a consequence, algorithms based on statistical modelling may not perform well in classification of hyperspectral data in full feature space [6]. On the contrary, among the machine learning algorithms, Support Vectors Machine (SVM) [7] have demonstrated outstanding abilities to deal with hyperspectral data [8], even in the situation of very small training sets [9]. However, standard SVM are pixel-based classifiers and no spatial information is used in the classification process. Recent studies have investigated different approaches to deal conjointly with the spatial and spectral information using SVM. For instance, Camps-Valls proposes to estimate the local mean and variance for each pixel using a fixed square neighborhood. These statistical parameters are included in the classification process using a composite kernel [10]. In [11], Mercier uses the same kernel formulation but the spatial information is estimated by a wavelets decomposition of the image. In this paper, we have adopted the same kernel strategy: a linear combination of two kernels is used, one acting the spectral information and one acting on the spatial information.

The originally of our proposed method is the definition of an *adaptive neighborhood*. This methodology is detailed in Section 2.1. Then, following [10, 11], we present the classification using the spatial and the spectral information with the SVM based on a mixture of kernels, Section 3. Experimental results are presented and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. CONTEXTUAL INFORMATION

In this section, we present how neighbors-sets are defined and how the contextual information is estimated.

The authors would like to thank Professor David A. Landgrebe for providing the hyperspectral data and the reference data used in these experiments. This research was supported in part by the Research Fund of the University of Iceland and the Jules Verne Program of the French and Icelandic governments (PAI EGIDE).

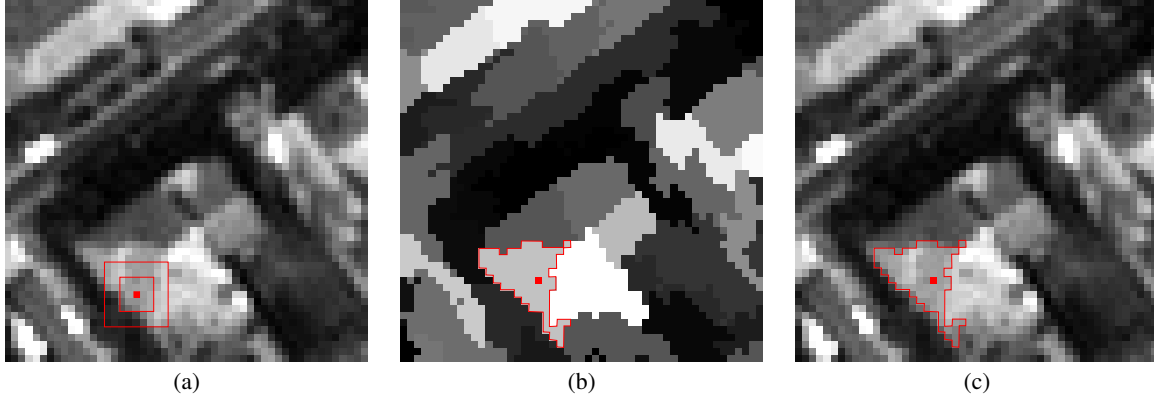


Fig. 1. Inter-pixel dependency estimation. (a) Original image and fixed square neighborhood. (b) Filtered image and neighbors-set defined using flat zones. (c) Original image with the defined neighbors-set Ω_x .

2.1. Self-complementary area filters

Self-complementarity is a very important property: the operator is its own complement ($\Psi = \mathbf{C}\Psi$), which is stronger than self-duality ($\Psi = \mathbf{C}\Psi\mathbf{C}$) [12]. A self-complementarity filter treats structures in the image that have different local contrast equally, which is not possible with the self-duality property only. Thus, they analyze all the structures of an image at once, local *extrema* (be they *minima* or *maxima*) as well as regions with intermediate grey levels.

Self-complementary area filters have been introduced to extend area opening and closing to all the structures of the image, not only its local *extrema* [12]. This is of the utmost interest for the analysis of very high resolution remote sensing images: these connected filters enable to remove small meaningless structures (*e.g.* cars on the road) while preserving borders of interest with a very high accuracy and independently of the structures' contrast. P. Soille has proposed a two steps algorithm, which consists of:

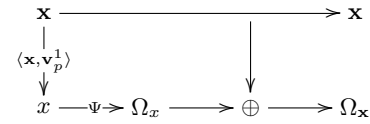
1. Labelling all the flat zones that satisfy the area criterion λ ,
2. Growing the labelled flat zones until a partition of the image is reached.

A better preservation of the image structures is achieved by iterating the algorithm until the desired minimal size is obtained, *e.g.*, let I be the image to process, then $\Psi_\lambda(I) = \Psi_\lambda(\Psi_{\lambda-1}(\dots(\Psi_2(I))))$, where λ is the minimal size of the remaining flat structures.

As stated in the introduction, we define the neighborhood of each pixel as the connected zone resulting from the application of a self-complementary area filter. This is illustrated in Fig. 1 where Fig. 1.(b) is the area filtering of Fig. 1.(a). The image is partitioned into flat zones. Each flat zone is consistent and hence belongs to one single structure in the original image. Furthermore, the smallest structures have been removed and only the main structure of interest remain. All the pixels belonging to one given flat zone are considered as neighbors. The neighborhoods defined in this way are applied on the original image. Fig. 1.(c) shows the morphological neighborhood Ω_x associated with the spotted pixel x . It is obviously more homogeneous and spectrally consistent than the fixed square featured on Fig. 1.(a).

The direct use of the area filter to hyperspectral remote sensing images is not possible, because of the lack of ordering relation. In order to overcome this shortcoming, several approaches can be considered. Using marginal ordering, one can apply the area filter on each band independently, but considering the high inter-band correlation, this is not appropriate [13, 14]. In this article, we adopt

the same strategy than in [15], where a Principal Component Analysis (PCA) [5] is applied to reduce the dimensionality and project data onto vector space where ordering relation exist. Then we compute the area filtering on the first principal component to extract the neighborhood of each pixel. The neighbors mask is applied on each band of the data and spatial information are extracted. The following scheme sum up the methodology:



where \mathbf{v}_p^1 is the first eigenvector corresponding to the largest eigenvalue.

2.2. Extracting spatial feature

Once the neighborhood of each pixel has been adaptively defined, spatial information is extracted. Considering the small average size of the neighbors set, a description using order statistics would not be reliable. Shape descriptors are not appropriate either as one given structure might be split into several consistent regions. (see Fig. 1.(b): The roof is divided into several triangles). As a conclusion, we propose to compute the vector median value of the neighbors set Ω_x , for every pixel x :

$$\Upsilon_x = \text{med}(\Omega_x) \quad (1)$$

where $\dim(\mathbf{x}) = \dim(\Upsilon_x) = n$, the number of spectral bands. Unlike the mean vector, the median vector is a vector from the initial set, which ensures a certain spectral consistency.

As a conclusion, every pixel now has two features: The *spectral feature* \mathbf{x} which is the original value of each pixel and the *spatial feature* Υ_x which is the median value computed on each pixel's adaptive neighborhood. The easiest way to use both information would be to build a stack vector. In this paper, we propose to exploit the kernel trick [16] of the SVM and design a composite kernel that allows to tune the relative influence of the extracted features. This is detailed in the next section.

3. SUPPORT VECTORS MACHINES (SVM)

The SVM classifier is surely one of the most popular kernel learning algorithms. It performs robust non-linear classification using the kernel trick. The idea is to find a separating hyperplane in some feature space induced by the kernel function while all the computations are done in the original space [16]. Given a training set $\mathcal{S} = \{(\mathbf{x}^1, y_1), \dots, (\mathbf{x}^\ell, y_\ell)\} \in \mathbb{R}^n \times \{-1; 1\}$, the decision function is found by solving the convex optimization problem:

$$\begin{aligned} \max_{\alpha} g(\alpha) &= \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{\ell} \alpha_i \alpha_j y_i y_j k(\mathbf{x}^i, \mathbf{x}^j) \\ \text{subject to} & \quad 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{\ell} \alpha_i y_i = 0 \end{aligned} \quad (2)$$

where α are the Lagrange coefficients, C a positive constant that is used to penalize the training errors, and k the kernel. To be an acceptable kernel, k should be a positive semi-definite function [17]. A short comparison of kernels for remotely sensed image classification can be found in [9].

When the optimal solution of (2) is found, *i.e.* the α_i , the classification of a sample \mathbf{x} is achieved by looking to which side of the hyperplane it belongs:

$$y = \text{sgn} \left(\sum_{i=1}^{\ell} \alpha_i y_i k(\mathbf{x}^i, \mathbf{x}) + b \right). \quad (3)$$

To construct the kernel that works with both the spatial and the spectral information at the same time, we use the positive linear combination of two kernels, namely the *spectro-spatial kernel*:

$$\mathcal{K}_{\sigma, \mu}^{\lambda}(\mathbf{x}^i, \mathbf{x}^j) = \mu k_{\sigma}^{\text{spect}}(\mathbf{x}^i, \mathbf{x}^j) + (1 - \mu) k_{\sigma}^{\text{spat}}(\mathbf{x}^i, \mathbf{x}^j) \quad (4)$$

where λ is the area filtering parameter and μ a weighting parameter which controls the relative influence of each kernel. The latter parameter μ has to be tuned during the training process. The spectral kernel $k_{\sigma}^{\text{spect}}$ is chosen as the standard Gaussian kernel:

$$k_{\sigma}^{\text{spect}}(\mathbf{x}^i, \mathbf{x}^j) = \exp \left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|^2}{2\sigma^2} \right) \quad (5)$$

where the norm is the *Euclidean*-norm and $\sigma \in \mathbb{R}^+$ tunes the flexibility of the kernel. The spatial kernel k_{σ}^{spat} is also Gaussian but uses the extracted spatial features $\Upsilon_{\mathbf{x}}$:

$$k_{\sigma}^{\text{spat}}(\mathbf{x}^i, \mathbf{x}^j) = \exp \left(-\frac{\|\Upsilon_{\mathbf{x}^i} - \Upsilon_{\mathbf{x}^j}\|^2}{2\sigma^2} \right). \quad (6)$$

The *one versus all* strategy was chosen for the multi-class classification problem [7]. It allows us to tune the parameter μ for each class independently.

4. EXPERIMENTAL RESULTS

Hyperspectral data from the optical AVIRIS sensor were used in the experiments. The data contains 220 data channels and is 145×145 pixels. It was collected in June 1992 over the Indian Pine Test site. Sixteen classes of interest are defined (see Table 1). The reference data consist of 10366 pixels that are not uniformly distributed; class 9 has only 20 labelled pixels while class 11 has 2468 labelled pixels. All the spectral bands were used, even though some of them were known to be noisy. No pre-processing were done.

Different training sets were randomly constructed from the reference data with 5, 10, 15, and 20 pixels by class, respectively. Each

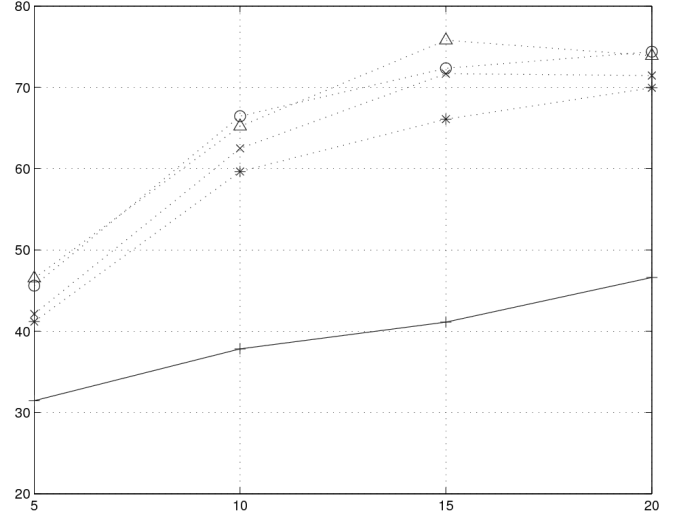


Fig. 2. Kappa coefficient of agreement. The horizontal axis represents the number of training samples use for each class and the vertical axis represents the κ values in percentage (+ : Classical SVM using the spectral information only; * : proposed approach with $\lambda = 5$; x : Proposed approach with $\lambda = 9$; o : Proposed approach with $\lambda = 13$; Δ : Proposed approach with $\lambda = 17$).

experiment is repeated five times and the averaged results are reported. The validation was done using the whole reference data.

The accuracy is assessed using the Kappa coefficient of agreement κ , *i.e.*, the percentage of agreement corrected by the amount of agreement that could be expected due to chance alone [18].

Several sizes of the area filtering parameter λ were tested. According to the low spatial resolution, we have used relatively low spatial parameter: from 3 to 20 pixels. The SVM parameters were tuned for each binary sub-problem by cross-validation: $\sigma^2 = \{0.5, 1, 2, 4\}$, $\mu \in \{0.1, 0.2, \dots, 0.9\}$ and $C = 200$. The optimization problem was solved using a modified version of the LIBSVM library (<http://www.csie.ntu.edu.tw/~cjlin/libsvm>).

Experimental results are given in Fig. 2. With 5 pixels by classes, corresponding to less than 1% of the available reference data, the κ is as low as 31.4 using the spectral information only. On the contrary, a significant increase of κ is obtained when using the additional spatial information with the proposed method (typically around 50% increase), for instance $\kappa = 47.30$ for $\lambda = 17$. Naturally, better classification accuracies are obtained using the larger training sets. For instance, when 20 training samples are used for each class, κ reaches over 70 whatever the value of λ , while using only the spectral bands leads to $\kappa = 46.6$.

Note that when considering larger area thresholds ($\lambda > 20$), the classification accuracy decreases since the area filter is too strong and structures of interest start to disappear. Consequently, one should set λ according to the spatial resolution of the data and the minimal size of the structures one is interested in.

Table 1 sums up the kernel parameters found during the training process for each binary sub-problem. The value of μ differs for the different classes. This tends to prove that the amount of each kind of information actually needs to be tuned carefully during the training process and should not be fixed *a priori* to the same value for every class.

The obtained results are very promising. They clearly outper-

Table 1. Kernel parameters found for $\lambda = 17$ and 20 training samples by classes. $\kappa=75.80$, Overall Accuracy: 78.78%, Average Accuracy: 85.57%.

Classes	1	2	3	4	5	6	7	8
Name	Alfalfa	Corn-notill	Corn-min	Corn	Grass/Pasture	Grass/Tress	Grass/Pasture-mowed	Hay-windrowed
μ	0.6	0.4	0.5	0.4	0.4	0.1	0.9	0.4
σ^2	0.5	0.5	0.1	4	0.5	2	4	2
Classes	9	10	11	12	13	14	15	16
Name	Oats	Soybeans-notill	Soybeans-min	Soybeans-clean	Wheats	Woods	Bldg-Grass-Tree-Drives	Stone-steel Towers
μ	0.7	0.1	0.1	0.2	0.7	0.3	0.3	0.8
σ^2	4	2	2	1	0.5	0.5	4	0.5

form results obtained on the same data set and presented in [1], pages 334-346. With supervised feature reduction (Decision Boundary Feature Extraction [1]) and a Gaussian Maximum Likelihood classifier, the overall accuracy was only 63.3% for a training set made of 688 samples and 72.4% for a training set made of 1490 samples. Using the same reference set, we achieved 78.78% of overall accuracy, with only 320 training samples see Table 1. Also, substantial improvement over the classical SVM is obtained, as can be seen from Fig. 2.

5. CONCLUSION

The classification of hyperspectral data with a spectro-spatial SVM classifier is investigated. Influence of a very small training set is studied in an empirical way. In order to include spatial information in the analysis, an adaptive definition of every pixel's neighborhood is proposed. Using an area self-complementary filter, the image is simplified to remove irrelevant structures. The remaining flat zones are considered as sets of neighbors. For each flat zone, the vector median value is extracted to take spatial neighborhood into account. The classification is performed using support vector machines and a spectro-spatial kernel. Comparisons with state of the art algorithms, a Gaussian ML and a SVM, showed a clear increase of the classification accuracy.

Some open questions remain to be addressed. The parameter λ in the area filtering has to be tuned. It does not seem critical, as there is a large range of values for which the classification accuracy remains almost equal. Furthermore, prior knowledge on the data resolution and the size of the structures of interest can be used. The kernel parameter has also to be tuned properly, some *a priori* information on the physical nature of the class should help to reduce the range of tested values and thus reduce the global training time. Our ongoing works include the definition of additional spatial information, such as textural information.

6. REFERENCES

- [1] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley and Sons, New Jersey, 2003.
- [2] Q. Jackson and D. A. Landgrebe, "Adaptive bayesian contextual classification based on markov random field," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, pp. 2454–2463, Nov. 2002.
- [3] S. Le Hégat, A. Kallel, and X. Descombes, "Ant colony optimization for image regularization based on a non-stationary markov modeling," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 865–878, Mar. 2007.
- [4] D. L. Donoho, "High-dimensional data analysis: the curses and blessing of dimensionality," in *AMS Mathematical challenges of the 21st century*, 2000.
- [5] K. Fukunaga, *Introduction to statistical pattern recognition*, CA: Academic Press, San Diego, 1990.
- [6] L. Jimenez and D. A. Landgrebe, "Supervised classification in high dimensional space: geometrical, statistical and asymptotical properties of multivariate data," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, no. 1, pp. 39–54, February 1993.
- [7] B. Schölkopf and A. J. Smola, *Learning with Kernels. Support vector machines, regularization optimization and beyond.*, MIT Press, 2002.
- [8] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, August 2004.
- [9] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP'06. Proceedings*, May 2006.
- [10] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.
- [11] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2839–2846, Oct. 2006.
- [12] P. Soille, "Beyond self-duality in morphological image analysis," *Image and Vision Computing*, vol. 23, no. 2, pp. 249–257, 2005.
- [13] J. Chanussot and P. Lambert, "Total ordering based on space filling curves for multivalued morphology," in *4th International Symposium on Mathematical Morphology and its Applications*, June 1998, pp. 51–58.
- [14] P. Lambert and J. Chanussot, "Extending mathematical morphology to color image processing," in *CGIP'00 - 1st International Conference on Color in Graphics and Image*, Oct. 2000, pp. 158–163.
- [15] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [16] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [17] N. Aronszajn, "Theory of reproducing kernel," *Tech. Rep. 11*, Harvard University, Division of engineering sciences, 1950.
- [18] C. Liua, P. Frazier, and L. Kumar, "Comparative assessment of the measures of thematic classification accuracy," *Remote Sensing of Environment*, vol. 107, no. 4, pp. 606–616, Apr. 2007.