CREATE THE RELEVANT SPATIAL FILTERBANK IN THE HYPERSPECTRAL JUNGLE

Devis Tuia[1*], *Michele Volpi*[2], *Mauro Dalla Mura*[3], *Alain Rakotomamonjy*[4], *Remi Flamary*[5]

[1] Ecole Polytechnique Fédérale de Lausanne (Switzerland) [2] University of Lausanne (Switzerland)
[3] Grenoble Institute of Technology (France) [4] University of Rouen (France)
[5] University Nice Sophia Antipolis (France)

ABSTRACT

Inclusion of spatial information is known to be beneficial to the classification of hyperspectral images. However, given the high dimensionality of the data, it is difficult to know before hand which are the bands to filter or what are the filters to be applied. In this paper, we propose an active set algorithm based on a l_1 support vector machine that explores the (possibily infinite) space of spatial filters and retrieve automatically the filters that maximize class separation. Experiments on hyperspectral imagery confirms the power of the method, that reaches state of the art performance with small feature sets generated automatically and without prior knowledge.

1. INTRODUCTION

When dealing with hyperspectral image (HSI) classification, a crucial issue is the inclusion of spatial information [1], especially when the spatial resolution is high. Many papers have stated the importance of including information about the spatial context of the pixels, either in the form of additional features in the input space [2, 3], regularizers in the classifier [1, 4] or through segmentation [5]. In this paper, we consider the first approach, i.e. the enrichment of the spectral information by inclusion of spatial features.

The problem of defining a relevant (i.e., discriminant) contextual feature set for remote sensing image classification is difficult: there are many contextual feature types (ex: textural, morphological, attribute filters, Gabor, wavelets, ...), each one with several filters (taking textural features as an example, we can have occurrence and co-occurrence features, as well as many filters such as average, variance, entropy, ...) and parameters (shape and size of the sliding window, orientation, level of decomposition, ...). The search space is thus very large (possibly infinite in the case of continuous-valued functions such as wavelets or Gabor filters) and selecting the good filters/parameters beforehand is complex. In remote sensing, the problem is usually circumvented by generating a large contextual filterbank driven by the user's prior knowledge on the problem. The enriched feature space is used to build a robust classifier and then some dimensionality reduction is performed aiming at extracting the relevant features: in [2], the authors use discriminant linear feature extraction, in [6] authors prune a neural network trained on the whole filterbank and finally, in [7] the authors use a recursive algorithm based on the SVM decision function to retrieve the most important features. Even if successful, all these algorithms show as drawback the need for the prior definition of the filterbank containing the relevant features.

This limitation becomes even more constraining when considering HSI, for which the dimensionality of the input space (spectral only) can reach a few hundreds. In this already large input space, it becomes unfeasible to generate the complete filterbank covering all possible filters computed over all bands. In remote sensing literature, this is usually dealt with by computing the filterbank on the first PCA(s) extracted from the HSI [2]. However, passing through a feature extraction step reduces the information and the relevant features can (again) be missed. Moreover, the physical meaning of the HSI bands is lost, while a subset of the original features can still reveal properties of the surveyed surfaces [4].

In this paper, we propose an effective way to tackle both problems simultaneously in an efficient (the model is linear), compact (the model is sparse) and exploratory (no restrictions on the type of filter or the bands to filter) way. By exploiting the properties of the l_1 regularized SVM, we define a large margin fitness function and an active set algorithm to search in the space of all possible filters and parameters and *discover* those that would increase the margin if added to the current input space [8]. The proposed approach proceeds iteratively by exploring the (potentially infinite) space of parameters of spatial filters in order to retrieve the features that are the most promising for the classification. The search is performed separately for each class (although this is not a specific requirement of the method) to allow each class to retrieve the features that best adapt to its spatial properties.

We test the proposed algorithm on a HSI acquired by the ProSpecTIR sensor with 360 spectral bands. Numerical results are competitive with respect to l_2 algorithms using a predefined complete filterbank and also select efficiently the relevant features for each class in a compact way.

* devis.tuia@epfl.ch

This work was partly funded by the Swiss National Science Foundation under the grants PZ00P2-136827 and 200020-144135 and by the French ANR 09-EMER-001 and ANR-12-BS03-003. The author would like to thank Prof. M. M. Crawford for the access to the PROSpecTIR data.

2. METHOD

Consider a binary classification problem with n training examples $\{\mathbf{x}_i, y_i\}_{i=1}^n$, where $\mathbf{x}_i$ is the feature vector $\in \mathbb{R}^b$ (a pixel in a b-dimensional image) and $y_i \in \{-1, 1\}$ is its label. We define a θ-parametrized function $\phi_\theta(\cdot)$ that maps a given pixel into his feature space (the output of a spatial filter). In this framework, we are looking for a decision function of the form $f(\mathbf{x}) = \sum_{j=1}^d w_j \phi_{\theta_j}(\mathbf{x})$, with $\mathbf{w} = [w_1, \ldots, w_d]^T$ the vector of all weights in the decision function. Note that this function considers only a finite number of feature maps d with associated parameters $\{\theta_j\}_{j=1}^d$. We define $\mathbf{\Phi}_{\theta_j}$ as the vector whose rows i are $\phi_{\theta_j}(\mathbf{x}_i)$ and $\mathbf{\Phi}$ as the $(n \times d)$ matrix of feature maps, resulting from the concatenation of the d vectors $\{\mathbf{\Phi}_{\theta_j}\}$. Each column of $\mathbf{\Phi}$ is normalized to unit norm. We also define $\widetilde{\mathbf{\Phi}} = \text{diag}(\mathbf{y})\mathbf{\Phi}$, with $\mathbf{y}$ being the vector of labels $\{y_i\}$. We learn the f function by optimizing the following ℓ_1 regularized linear SVM problem:

$$\min_{\mathbf{w}} \quad \frac{1}{2}(\mathbb{I} - \widetilde{\mathbf{\Phi}}\mathbf{w})_+^T(\mathbb{I} - \widetilde{\mathbf{\Phi}}\mathbf{w})_+ + \lambda\|\mathbf{w}\|_1 \qquad (1)$$

where $[\widetilde{\mathbf{\Phi}}\mathbf{w}]_i = y_i f(\mathbf{x}_i)$, $\mathbb{I}$ is a vector of ones, $(\cdot)_+ = \max(0, \cdot)$ is the element-wise positive part of a vector and λ is the regularization parameter. Note that the left term in Eq. (1) is the differentiable squared hinge loss. The optimality conditions of this problem [9] are:

$$r_{\theta_j} + \lambda \text{sign}(w_i) = 0 \qquad \forall j \quad w_j \neq 0 \qquad (2)$$

$$|r_{\theta_j}| \leq \lambda \qquad \forall i \quad w_j = 0 \qquad (3)$$

with $r_{\theta_j} = \widetilde{\mathbf{\Phi}}_{\theta_j}^T(\mathbb{I} - \widetilde{\mathbf{\Phi}}\mathbf{w})_+$ the scalar product between $\widetilde{\mathbf{\Phi}}_{\theta_j}$ and the hinge loss error. This means that at optimality $|r_{\theta_j}| \leq \lambda \; \forall\theta_j \in \varphi$. If we extend this reasoning a little further we can also assume that, for the current model, $|r_{\theta_j}| \leq \lambda \; \forall\theta_j \notin \varphi$: this means that with a fixed set of active filters φ all the other possible filters receive a null weight in the current model. If this assumption holds, any new feature violating such constraint, i.e. any feature $\phi_{\theta_j}^* \notin \varphi$ with corresponding $|r_{\theta_j}^*| > \lambda$, will lead to a decrease to the objective function if added to φ.

This leads to the development of an active set algorithm that solves iteratively Eq. (1), restricted to the features in the current active set. At each iteration, if a feature not in the active set (*i.e.* $w_j = 0$) violates optimality constraint (3), it is added to the active set of the next iteration, yielding a decrease of the objective value after re-optimization.

By iteratively adding single features to the current active set, we perform a search in the space of possible features, and simultaneously optimize a max-margin classifier. The algorithm has linear $\mathcal{O}(n)$ complexity for a given iteration.

When dealing with continuously parametrized features, the number of candidate features to be screened becomes possibly infinite, so an exhaustive test of all the candidate features is intractable. To cope with this problem, we generate random subsets of possible features with vectors $\{\theta_j\}_{j=1}^p$ selected in the set of possible values Θ (in the experiments reported in this paper, such set is detailed in Table 1).

The random filters generator draws a vector or parameters $\boldsymbol{\theta} = [\theta_1, ..., \theta_j, ..., \theta_p]$, where each θ vector contains a band identifier, a filter family, type and parameters. The filters are generated and r_{θ_j} are calculated with the model using the current active set. The feature $\phi_{\theta_j}^*$ mostly violating the constraint in Eq. (3) is added to the active set and the process is iterated. The iterative procedure is detailed in Algorithm 1. The algorithm stops when a stopping criterion, such as a maximum number of iterations, a sequence of filterbanks realizations without violating features or a threshold on the decrease in the SVM objective function, is met.

3. RESULTS

3.1. Setup

We tested the active set algorithm on a HSI image acquired by the ProSpecTIR system near Purdue University, Indiana, on May 24-25, 2010. The image contains 445×750 pixels at 2-m spatial resolution, with 360 spectral bands of 5-nm width. Sixteen land cover classes were identified, which included fields of different crop residue covers, vegetated areas, and man-made structures. Many classes have regular geometry associated with fields, while others are associated with roads and isolated man-made structures. The two first columns of Fig. 1 illustrate the image in a RGB composition and the available ground reference.

We used 100 labeled pixels per class to optimize the linear one against all SVM. We report average results over five independent starting training sets at the end of the process, as well as the respective standard deviations. We run the active set algorithm for 200 iterations for each class, thus discovering

Algorithm 1 Active set algorithm

Inputs
- Initial active set φ (bands or PCAs)
- Filters characteristics Θ (Table 1)
- Tolerance on the optimality constraint violation : ϵ

repeat
  Solve l_1 SVM with current active set φ
  if no features selected at previous iteration **then**
    Generate a new filterbank $\{\phi_{\theta_j}\}_{j=1}^p \notin \varphi$
  end if
  for j = 1: p **do**
    Compute $r_{\theta_j} = \widetilde{\mathbf{\Phi}}_{\theta_j}^T(\mathbb{I} - \widetilde{\mathbf{\Phi}}\mathbf{w})_+$
  end for
  Find feature $\phi_{\theta_j}^*$ maximizing $|r_{\theta_j}|$
  If $|r_{j^*}| > \lambda + \epsilon$, add the feature to the active set $\varphi = \phi_{\theta_j}^* \cup \varphi$
until stopping criterion is met

Table 1. Filters used in the experiments, along with their parameters and possible values

Bank	Filters	Parameters	Type	Search range
All filters		- Band (or PCA)	int	$[1 : b]$
Morphological (`MOR` [2])	Opening, Closing, Opening top-hat, Closing top-hat, Opening by reconstruction, Closing by reconstruction, Opening by reconstruction top-hat and Closing by reconstruction top-hat	- Shape of structuring element	str	{disk, diamond, square, line}
		- Size of structuring element	int	$[1 : 15]$
		- Angle (if Shape = 'line')	float	$[0, \pi]$
Texture [6]	Mean, Range, Entropy and Std. dev.	- moving window size	int	$[5 : 2 : 21]$
Attribute (`ATT` [10])	Area Diagonal	- Area - Diagonal of bounding box	int int	[100, 10000] [10, 100]

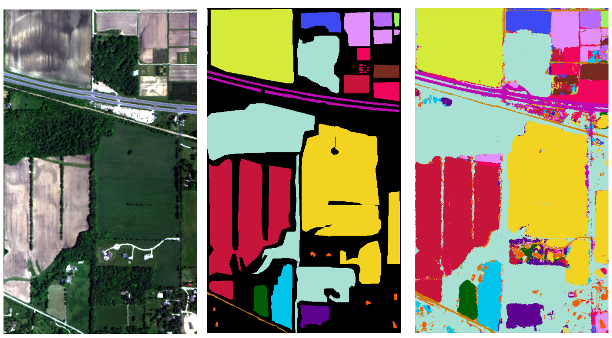

image ground truth classification

Fig. 1. Left: ProSpecTIR image in RGB composition; middle: ground reference (16 classes); right: classification map obtained with the proposed active set algorithm (one run of the `AS-PCAs` setting).

the relevant features for each one separately.

We consider two settings: the first filters the b original bands (`AS-Bands`), while the second uses the first 50 PCA projections as base images (`AS-PCAs`). For each experiment, the spatial filterbank contains three features types, namely texture, morphological and attribute filters. The range of possible filters and parameters is reported in Table 1.

3.2. Results

Numerical performances of the proposed method are reported in Table 2: results of the active set algorithm (`AS-Bands` and `AS-PCAs`) are reported in the four top rows, while in the rest we report performances obtained using SVM and pre-defined input spaces: the original bands (`Bands`), the 10 first PCAs (`PCA`), the ensemble of possible morphological filters with parameters given in Table 1 (`MOR`) and the same for attribute filters (`ATT`) and the totality of filterbanks in the table (`ALL`). We compared the proposed method against i) a l_1 SVM performing simultaneously selection and classification, ii) a l_2 SVM trained on the features selected by the l_1 model and iii) a l_2 SVM trained on all the features (without selection). Note that the l_2 model trained on the selected features is also considered for the proposed method.

The proposed method performs remarkably well, with average classification accuracy between 96.7% (when using directly the l_1 classifier) and 99.3% (when retraining a l_2 SVM on the selected features). These results are obtained without prior knowledge of the relevant features and without pregenerating the entire filterbanks. They are superior to almost all the other experiments, with the exception of the l_1 experiment using the `ALL` input features generated on the three first PCAs. The classification map obtained by the `AS-PCAs` method is reported in the right column of Fig. 1.

The proposed method also returns filterbanks that are much more compact than the competing methods: the `AS-` methods select on average a total of 369 features to solve the problem, which is half of the size of the most compact set retrieved by all the other methods, with the exception of the `PCA` setting, which is sparser, but also related to a 5%-7% lower classification performance. Enforcing more sparsity in the l_1 classifier with pre-defined sets (by increasing the λ parameter) returns much compact feature sets, but at the cost of heavily degraded classification performance (losses between 10% and 25%, results not reported).

4. CONCLUSION

In this paper, we proposed a methodology to *discover* the important contextual features among a possibly infinite set of existing filters. Without imposing the nature and parameters of the filters in advance, we let the max-margin criterion select which filters would be useful for classification. Moreover, we perform selection separately for each binary classifier, thus allowing each class to add to its active set only the filters that are the most relevant for its spatial structures.

Results on a hyperspectral image illustrated the potential of the method, the obtained results are comparable to the l_2 counterpart, but with much more compact filterbanks and without any a priori definition of the filters. The method can also be used as a simple feature selector, after which the selected features are used in a l_2 classifier. Experiments in this sense also showed a boost in the SVM performances.

Table 2. Results of the proposed active set algorithm using original bands (AS-Bands) or the 50 first PCAs (AS-PCA). Results are compared to l_1 and l_2 SVMs using the original bands (Bands, no spatial information), the ten first PCAs (PCA, no spatial information) and contextual filters generated from the 3 first PCAs and the whole set of possible features in Table 1 (the ALL set contains all morphological, attribute and texture filters).

		Feature set	# input features Per class	Total	Active features (total) μ	σ	OA μ	OA σ	Kappa μ	Kappa σ
Active set	l_1	AS-Bands	*	*	369.40	13.11	93.57	2.74	0.922	0.033
		AS-PCAs	*	*	350.20	7.56	**96.72**	1.98	**0.960**	0.024
	l_2^+	AS-Bands	*	*	369.40	13.11	97.69	0.29	0.972	0.004
		AS-PCAs	*	*	350.20	7.56	**99.29**	0.22	**0.991**	0.003
Pre-generated fullbanks	l_1	Bands	360	5760	3186.20	43.10	89.15	0.53	0.869	0.006
		PCA (10 PCAs)	10	160	139.40	3.78	89.03	0.61	0.868	0.007
		MM (from 3 PCAs)	111	1776	686.00	17.97	92.09	0.44	0.904	0.005
		ATT (from 3 PCAs)	243	3888	1000.00	44.19	95.96	0.71	0.951	0.009
		All (from 3 PCAs)	381	6096	1293.20	43.65	**99.09**	0.16	**0.989**	0.002
	l_2^*	Bands	360	5760	3186.20	43.10	93.55	0.45	0.922	0.005
		PCA (10 PCAs)	10	160	139.40	3.78	87.25	0.73	0.847	0.008
		MM (from 3 PCAs)	111	1776	686.00	17.97	92.39	0.53	0.908	0.006
		ATT (from 3 PCAs)	243	3888	1000.00	44.19	96.45	0.45	0.957	0.005
		All (from 3 PCAs)	381	6096	1293.20	43.65	99.01	0.26	0.988	0.003
	l_2	Bands	360	5760	5760.00	0.00	94.30	0.48	0.931	0.006
		PCA (10 PCAs)	10	160	160.00	0.00	87.32	0.73	0.847	0.009
		MOR (from 3 PCAs)	111	1776	1776.00	0.00	92.58	0.61	0.910	0.007
		ATT (from 3 PCAs)	243	3888	3888.00	0.00	93.04	0.41	0.915	0.005
		All (from 3 PCAs)	381	6096	6096.00	0.00	98.21	0.37	0.978	0.005

$^+$ = on features selected by the active set algorithm only

* = on features selected by the l_1 SVM only

5. REFERENCES

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1 –24, 2013.

[2] 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. 480–490, 2005.

[3] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804 – 3814, 2008.

[4] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780 – 3791, 2010.

[5] Y. Tarabalka, J. Chanussot, and J.A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Sys. Man Cybernetics B*, vol. 40, no. 5, pp. 1267 –1279, 2010.

[6] F. Pacifici, M. Chini, and W.J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, 2009.

[7] D. Tuia, F. Pacifici, M. Kanevski, and W.J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, 2009.

[8] A. Rakotomamonjy, R. Flamary, and F. Yger, "Learning with infinitely many features," *Machine Learning*, 2012.

[9] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Convex optimization with sparsity-inducing norms," in *Optimization for Machine Learning*. MIT Press, 2011.

[10] M. Dalla Mura, J. Atli Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, 2010.