Discovering relevant spatial filterbanks for VHR image classification

Devis Tuia
Lausanne EPFL, Switzerland
devis.tuia@epfl.ch

Michele Volpi
Université de Lausanne, Switzerland
michele.volpi@unil.ch

Mauro Dalla Mura
Fondazione Bruno Kessler, Italy
dallamura@fbk.eu

Rémi Flamary, Alain Rakotomamonjy
Université de Rouen, France
{*remi.flamary, alain.rakoto*}*@insa-rouen.fr*

Abstract

In very high resolution (VHR) image classification it is common to use spatial filters to enhance the discrimination among landuses related to similar spectral properties but different spatial characteristics. However, the filters types that can be used are numerous (e.g. textural, morphological, Gabor, wavelets, etc.) and the user must pre-select a family of features, as well as their specific parameters. This results in features spaces that are high dimensional and redundant, thus requiring long and suboptimal feature selection phases. In this paper, we propose to discover the relevant filters as well as their parameters with a sparsity promoting regularization and an active set algorithm that iteratively adds to the model the most promising features. This way, we explore the filters/parameters input space efficiently (which is infinitely large for continuous parameters) and construct the optimal filterbank for classification without any other information than the types of filters to be used.

1 Introduction

Recent advances in optical remote sensing opened new highways for spatial analysis and geographical applications. Urban planning, crops monitoring, tracking: all these applications are nowadays eased by the use of satellite images that provide a large scale and non-intrusive observation of the surface of the Earth.

One of the factors raising the interest of applicative communities for remote sensing is the advent of very high resolution ($\leq$ 5m, VHR) imagery. With such imagery, it is now possible to observe fine scale phenomena and to delineate objects in a very precise fashion. However, such level of detail comes with a price: the resolution is so high, that the spectral signature of landuse types becomes mixed, due to the presence of a plethora of objects on the ground that are visible in the image. This increases the landuse intraclass variance. For example, with a submetric resolution, the average signature of the tiles covering a roof can be contaminated by the presence of different objects such as, chimneys, windows, solar panels, etc.

To exploit VHR imagery efficiently, the remote sensing community has turned to machine learning and image processing techniques [1]. For the former, powerful classifiers have been designed, among which the Support Vector Machines (SVM) are probably the most successful [2]. For the latter, a series of spatial filters accounting for texture or shape of objects is often used to cope with the spectral variance introduced by VHR imaging. Among the 2-D filters proposed, texture [3] and mathematical morphology [4] are the most used.

A problem that often arises is the choice of the spatial filters: these filters are numerous and come with a wide range of parameters. Even if SVM can cope with high dimensional spaces better than other methods, the choice of the filterbank strongly affects the results, since it defines the input space in which examples are discriminated. In remote sensing literature, most of the proposed approaches to VHR image classification (if not all of them) first define a large filterbank with prior knowledge of the user and then runs a model, with a feature selection procedure [3, 4, 5, 6, 7]. If the relevant features are included in the filterbank, the procedure provides an efficient and compact model. But what if the relevant features and related parameters are *not* known beforehand?

In this paper, we consider the problem of *discovering the relevant filterbank* by an active set method based on SVM optimality conditions [8]. Contrarily to [5], which also proposes sparse feature representation, the feature filterbank is not fixed prior to analysis. The fitness is based on the hinge loss, thus relying on a large margin criterion. In our approach, we generate iteratively random filterbanks of known filter types and assess whether one of the new features would be useful if added to the model. This way, the high dimensional (and continuous) space of features is explored and the optimal set of filters for classification is retrieved. Experiments on a VHR image confirm the hypothesis.

2 Selecting from an infinite set of features

Consider a set of n training examples $\{\mathbf{x}_i, y_i\}_{i=1}^n$ where $\mathbf{x}_i$ corresponds to a pixel in the image and $y_i \in \{-1, 1\}$ to 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 whoe rows i are $\phi_{\theta_j}(\mathbf{x}_i)$ and $\mathbf{\Phi}$ as the 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 and $\widetilde{\mathbf{\Phi}} = \text{diag}(\mathbf{y})\mathbf{\Phi}$, 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:

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

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

$$r_i = 1 \qquad \forall i \quad w_i \neq 0 \qquad (2)$$
$$|r_i| \leq 1 \qquad \forall i \quad w_i = 0 \qquad (3)$$

with $r_i = \frac{C}{n}\widetilde{\mathbf{\Phi}}_i^T(\mathbb{I} - \widetilde{\mathbf{\Phi}}\mathbf{w})_+$ the scalar product between $\widetilde{\mathbf{\Phi}}_i$ and the hinge loss error. These optimality conditions suggest the use 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.* $w_i = 0$) violates optimality constraint (3), it is added to the active set of the next iteration, leading to a decrease of the cost after re-optimization. With continuously parametrized features, the number of candidate features is possibly infinite, so a comprehensive test of the candidate features is intractable. In this situation, [8] proposed to randomly sample a finite number of features and add to the active set the one violating the most constraint (3).

Note that the algorithm is designed to handle large scale datasets. Indeed checking the optimality conditions and selecting a new feature has complexity $\mathcal{O}(n)$ and solving the inner problem is performed only on a small number of features using an accelerated gradient algorithm (see [8]).

3 Data and experimental setup

Analyses have been carried out on a multi-spectral QuickBird dataset of a residential neighborhood of Zurich, Switzerland. The image was acquired in 2002 and it is composed by 4 bands of 329×347 pixels, with a geometrical resolution of 2.4 m. Figure 1 illustrates the original image and the associated ground truth. Data have been scaled by dividing the pixel vectors by the maximum value of the original 11-bit image. Random subsets composed by 5% of the comprehensive labeled set (2040 pixels out of the 40762 available) are extracted for the experiments. The test set is composed by the remaining 38722 examples. For classification, we used a linear ℓ_1-norm SVM, with regularization parameter C optimized by cross-validation to a value of 100. The same C value has been used for the proposed method.

For the ℓ_1-norm SVM, we pre-computed a filterbank composed by all the spectral bands and contextual filters extracted for each band separately. Selected filters are morphological operators (MOR) with disk-shaped structuring element, attribute filters (ATT) [10] and occurrence texture indicators. The parameters used for computing the features are sampled at uniform intervals from the ranges given in Tab. 1. Three experiments are reported: a stack of the original bands with the MOR features, a stack with the ATT features and a case involving all the three types. For the dimensionality of these sets, please refer to line '# Features' of Tab. 2.

For the proposed infinite active set scheme, we generate random filterbanks iteratively, using the codebook described in Tab. 1. At each iteration, a random subset of features is computed and the one most violating the constraints is retained. Note that the algorithm re-

Image (R-G-B) Ground Truth

Figure 1. Datasets used, legend in Tab. 2.

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

Bank	Filters	Parameters	Type	Search range
All filters		- Band	int	$[1 : \#bands]$
Morphological (MOR [4])	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 [3]	Mean, Range, Entropy and Std. dev.	- moving window size	int	$[5 : 2 : 21]$
Attribute (ATT [10])	Area	- Area	int	[100, 10000]
	Diagonal	- Diagonal of bounding box	int	[10, 100]
	Inertia	- Moment of inertia	float	[0.1, 1]
	Standard deviation	- Standard deviation	float	[0.5, 50]

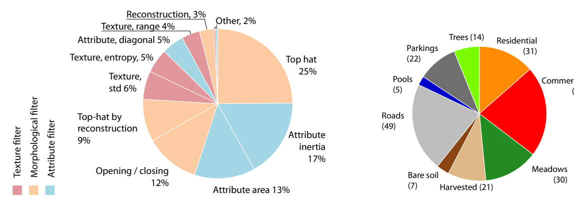
Table 2. Averaged numerical figures of merit of the strategies considered

Model	Predefined library				Infinite
	ℓ_1 SVM				
Feature type	Bands	MOR	ATT	All	
Overall accuracy	69.75	84.52	85.50	91.99	**92.46**
Cohen's Kappa	0.613	0.806	0.819	0.901	**0.907**
Residential	76.71	92.17	92.44	96.07	**96.71**
Commercial	51.49	74.02	66.42	79.65	**83.73**
Meadows	**99.93**	99.75	99.58	99.54	99.60
Harvested	0	30.47	83.24	**98.40**	97.51
Bare soil	49.53	**99.98**	99.41	99.93	99.91
Roads	88.92	84.50	84.32	88.95	**89.39**
Pools	21.09	95.47	**98.28**	97.42	96.40
Parkings	0	42.05	31.26	**56.41**	51.99
Trees	0	41.10	12.81	**65.98**	65.93
# Features	4	148	324	508	∞
# Selected	4	84.20	114.60	202.40	210.40

trieves class-specific filterbanks in the One Against All (OAA) architecture. For each class, we performed 200 generations of filters.

4 Results and discussion

Figures of merit are given in Tab. 2. Results reported are averages over 5 independent realizations of the training sets. The infinite feature learning algorithm provides the best results in terms of overall accuracy and estimated Cohen's kappa statistic [1], and ranks among the best for single class accuracy in most classes. The proposed active set method selects roughly the same number of features as the ℓ_1-norm SVM experiments, indicating a good convergence to the optimal situation in which filters are computed manually.

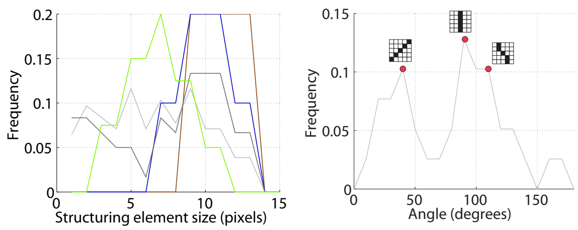


(a) (b)

Figure 2. Infinite active set algorithm: (a) selected filterbank per type and (b) number of retained features per class.

Figure 2(a) summarizes the number and type of filters selected by the proposed method. Features that model the geometry and morphology of regions (e.g., generated by attribute filters), as well as features in which locally dark/bright objects are extracted (i.e., by morphological top-hat operators) play an important role for this classification task (55% of the filterbank). This can be related to the nature of the classes to be detected, that ranges from structured geometry (roofs and roads) to large homogeneous patches (fields and meadows).

Regarding the class specific results, Fig. 2(b) illustrates the number of features required by each OAA subproblem. Classes related to large amounts of features are the 'road' and the 'commercial' classes, (47% of all the chosen features). From a spectral point of view these land use classes are ambiguous, since in the original input space they are not linearly separable from their counterparts 'parkings' and 'buildings'. Hence, more descriptors are needed to discriminate those classes. The same holds for the classes 'trees' and 'harvested', not spectrally separable from the class 'meadows' in the original spectral bands space. When using the spectral bands only, all the pixels are attributed to the most represented class ('meadows') and the other similar classes



(a) (b)

Figure 3. (a) Structuring element size within the morphological filters selected for five classes (for color legend, refer to Tab. 2). (b) Orientation of linear structuring elements for the class 'roads'.

are never predicted by the SVM (0% accuracy). On the contrary, the proposed active set method discovers features related (both in terms of types and quantities) to the degree of complexity of the class represented. To illustrate this principle even further, Fig. 3(b) reports the size of the structuring elements chosen for five classes: for spatially wide classes, the structuring element size is larger ('parkings', 'pools' and 'bare soil'), while for classes related to small or thin objects ('trees' and 'roads'), the size of the structuring elements are smaller. The example of the class 'roads' is also interesting, because of the geometrical properties of the class: since roads are linear structures, line structuring elements are often selected (26% of the features) and orientation of the filters correspond to the main street directions observed in the image (Fig. 3(b)).

5 Conclusions

In this paper, we presented and discussed a ℓ_1-norm SVM-based feature learning scheme for image classification. This scheme is based on active sets and discovers interesting features from a potentially infinite filter library. Without any prior knowledge other than filters types, the proposed method is able to identify the contextual information needed to correctly classify the OAA subproblems. This way, besides a high classification accuracy, the user is also provided with a filterbank describing the main spatial characteristic of the classes. It is also shown that not linearly separable classes become separable after the automatic dimensional expansion of the input space, producing accuracies close or even superior to a feature space created by an expert user. This opens new perspectives for VHR hyperspectral imaging, where the extraction of contextual filters in an input space of several hundreds of bands remains unsolved.

Acknowledgments

Authors acknowledge the support of the Swiss National Science Foundation (PZ00P2-136827, 200021-126505) and of the French ANR (09-EMER-001).

References

[1] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jimenez, and J. Malo. *Remote Sensing Image Processing*. Synthesis lectures on image, video, and multimedia processing. Morgan & Claypool, 2011.

[2] G. Camps-Valls and L. Bruzzone. *Kernel Methods for Remote Sensing Data Analysis*. J. Wiley & Sons, NJ, USA, 2009.

[3] 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.*, 113(6):1276–1292, 2009.

[4] 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.*, 43(3):480–490, 2005.

[5] K. Tieu and P. Viola. Boosting image retrieval. *Int. J. Computer Vision*, 56(1/2):17–36, 2004.

[6] R. Archibald and G. Fann. Feature selection and classification of hyperspectral images with support vector machines. *IEEE Geosci. Remote Sens. Lett.*, 4(4):674–679, 2007.

[7] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski. Learning relevant image features with multiple kernel classification. *IEEE Trans. Geosci. Remote Sens.*, 48(10):3780 – 3791, 2010.

[8] R. Flamary, F. Yger, and A. Rakotomamonjy. Selecting from an infinite set of features in SVM. In *European Symposium on Artificial Neural Networks ESANN*, 2011.

[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.*, 48(10):3747–3762, 2010.