HAL
open science

Learning spatial filters for multispectral image segmentation.

Devis Tuia, Rémi Flamary, Gustavo Camps-Valls, Alain Rakotomamonjy

To cite this version:

Devis Tuia, Rémi Flamary, Gustavo Camps-Valls, Alain Rakotomamonjy. Learning spatial filters for multispectral image segmentation.. Machine Learning for Signal Processing, Aug 2010, Kittila, Finland. pp.1-6. hal-00528923

HAL Id: hal-00528923

https://hal.science/hal-00528923v1

Submitted on 22 Oct 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

LEARNING SPATIAL FILTERS FOR MULTISPECTRAL IMAGE SEGMENTATION

Devis Tuia, Gustavo Camps-Valls

Image Processing Laboratory (IPL)
Universitat de València, Spain
{devis.tuia, gustavo.camps}@uv.es

Rémi Flamary, Alain Rakotomamonjy

Laboratoire LITIS
Université de Rouen, France
{remi.flamary, alain.rakoto}@insa-rouen.fr

ABSTRACT

We present a novel filtering method for multispectral satellite image classification. The proposed method learns a set of spatial filters that maximize class separability of binary support vector machine (SVM) through a gradient descent approach. Regularization issues are discussed in detail and a Frobenius-norm regularization is proposed to efficiently exclude uninformative filters coefficients. Experiments carried out on multiclass one-against-all classification and target detection show the capabilities of the learned spatial filters.

1. INTRODUCTION

In recent years, kernel methods and in particular support vector machines (SVM) [1, 2], have been extensively used for the classification (segmentation) of multispectral, hyperspectral and radar images with excellent results [3]. Among the problems raised by satellite imagery, one of the most challenging is the one regarding the increase of spatial resolution brought by last generation satellites, named Very High Resolution (VHR), that provide imagery at metric –or even submetric– resolution. Besides the obvious advantages of such increase in image resolution, VHR imagery raises a series of signal processing problems related, in particular, to the change in the spectral signature of the classes to be detected. The finer spatial resolution implies the increase of the intraclass variance of the classes, since small objects, such as chimneys or pedestrians, become visible in the image. Therefore, the signature of a particular class becomes highly variable and robust algorithms exploiting spatial information becomes strictly necessary. Certainly, such high geometrical resolution could be considered as a noise source but undoubtedly may also be beneficial because it allows a clearer definition of shapes and edges of the objects, especially in urban settings.

A great attention has been paid to *spatial filtering* of remote sensing images to improve recognition. Two basic families of spatio-spectral classifiers are encountered in the literature: (1) feature extraction (or preprocessing) techniques, and (2) filtering (or post-classification) methods. In the first case, the spectral signature of a given pixel is combined with the spatial information obtained from the neighbors through window or region-based approaches, such as morphological filtering [4, 5], geometrical features [6], or Markov random fields [7, 8]. The spatial and spectral feature vectors are then jointly classified. The second approach essentially performs spatial smoothing of a pixel-based classification map, and is often called *post-regularization*. The approach is mainly based on morphological operators, such as the majority voting scheme [9–12]. The generally low improvement in accuracy of these methods is due to the strong dependence of classifier's performance and filter design.

Considering spectral and spatial information jointly in the classifier typically leads to better results. However, extracting contextual features for classification gives rise to several problems: 1) critical parameters need to be tuned, such as the scale and extent of the spatial relations; 2) the quality of the feature extraction technique may hamper the representation capabilities of the classifier; and very importantly, 3) by combining spatial and spectral information sources, the *curse of dimensionality* problem is worsened. The latter problem was alleviated with the introduction of composite kernels [13], in which dedicated kernels for the spectral and spatial information are combined. The framework has been recently extended to deal with convex combinations of kernels through *multiple kernel learning* [14, 15]. In both cases, however, the methodology still relies on performing an *ad hoc* spatial feature extraction before kernel computation, typically limited to second-order statistics, Gray-level coocurrence matrices (GLCM) [16], or morphological reconstruction features [17, 18]. Yet successful, the approach strongly relies on user's previous knowledge about the particular image statistics to properly design the filters.

In this paper, we propose an alternative way to jointly *learn* (optimize) the SVM classifier and a convolution filter separating the class of interest from the others through

This work is supported by the Swiss National Foundation, Grant PBLAP2-127713, and by the Spanish Ministry of Innovation and Science under projects AYA2008-05965-C04-03 and CSD2007-00018.

a particular gradient-descent approach [19]. Since the preprocessing phase, i.e. the definition and selection of the filters, is integrated into the learning process, the proposed approach adapts the filter to the problem at hand, and is able to significantly improve classification results. The method is tested in two challenging VHR image classification scenarios: one-class target detection and multi-classification of land-cover types. The remainder of the paper is organized as follows: Section 2 describes the proposed algorithm and the optimization of the convolutional filter. Section 3 presents the data and the setup considered for the experiments discussed in Section 4. Finally, Section 5 concludes the paper.

2. LARGE MARGIN FILTERING FOR REMOTE SENSING IMAGES

In this section, large margin image filtering is proposed and issues regarding regularization are discussed in detail. In [19], the authors proposed to learn a one-dimensional convolution of the signals to enhance signal sequence labeling. In this paper, this approach is extended to pixel classification of remote sensing multidimensional images: in this case, the filtering corresponds to learning a convolution kernel that promotes maximum margin between classes. Additionally, the the filtered SVM proposed in [19] is here extended to the non-linear case, through Gaussian kernels.

2.1. Definitions

Consider an image $\mathbf{X} \in \mathbb{R}^{r_1 \times r_2 \times d}$, where $r_1 \times r_2$ is the spatial extent – the considered pixels – and d is the number of components, that can be spectral bands, textural features or any kind of index retrieved from the image. To ease notation, we define $\mathbf{X}_{i,j,k} = \mathbf{X}_{\mathbf{p},k}$ as the kth component of pixel $\mathbf{p} = (i, j)$. Since the approach proposed is supervised, a list of n labeled pixels $\mathbf{p} = (i, j) \in \mathcal{S}_l$ is available, as well as a corresponding labels vector $\mathbf{Y}_\mathbf{p} \in \{-1, 1\}, \forall \mathbf{p} \in \mathcal{S}_l$.

As stated in the introduction, a 2-dimensional convolution may be applied to the image in order to enhance discriminative information. In this case, the convolution is performed by applying per component convolution matrices that are stored in $\mathbf{F} \in \mathbb{R}^{f \times f \times d}$, where f is the size of the bidimensional square filters. When convoluting, the filter is centered on the current pixel. The filtered image $\widetilde{\mathbf{X}}$ is defined by:

$$\begin{aligned} \widetilde{\mathbf{X}}_{\mathbf{p},k} &= \sum_{u=1,v=1}^{f,f} \mathbf{F}_{u,v,k}\, \mathbf{X}_{i+u-f_0,j+v-f_0,k} \qquad (1) \\ &= \sum_{u=1,v=1}^{f,f} \mathbf{F}_{u,v,k}\, \mathbf{X}_{\mathbf{p}+(u,v)-f_0,k} \end{aligned}$$

where $f_0 = f/2$. The filter coefficients in $\mathbf{F}$ define the type of smoothing/enhancement performed by the convolution. For instance, smoothing can be achieved by applying an average filter corresponding to a matrix $\mathbf{F}_{i,j,k} = \frac{1}{f^2}, \forall (i, j, k)$.

Finally, we define the RBF kernel between the filtered pixels $\mathbf{p}$ and $\mathbf{q}$:

$$\widetilde{K}_{\mathbf{p},\mathbf{q}} = k(\widetilde{\mathbf{X}}_{\mathbf{p},.}, \widetilde{\mathbf{X}}_{\mathbf{q},.}) = \exp\left(-\frac{||\widetilde{\mathbf{X}}_{\mathbf{p},.} - \widetilde{\mathbf{X}}_{\mathbf{q},.}||^2}{2\sigma^2}\right), \quad (2)$$

where σ is the kernel width or bandwidth. This kernel will be used in all the following experiments.

2.2. Optimization problem

A possible approach to contextual filtering is to extract contextual filters and then select the relevant ones by feature selection. In this paper, we propose to jointly learn the pixel classifier and the 2D convolution with large margin constraints. This double objective can be formulated with the following minimization problem:

$$\min_{g,\mathbf{F}} \left\{ \frac{1}{2}\|g\|^2 + \frac{C}{n} \sum_{\mathbf{p}\in\mathcal{S}_l} H(\mathbf{Y}_\mathbf{p}, g(\widetilde{\mathbf{X}}_{\mathbf{p},.})) + \lambda\Omega(\mathbf{F}) \right\} \quad (3)$$

where $H(\mathbf{Y}_\mathbf{p}, g(\widetilde{\mathbf{X}}_{\mathbf{p},.}) = \max(0, 1 - \mathbf{Y}_\mathbf{p} \cdot g(\widetilde{\mathbf{X}}_{\mathbf{p},.}))$ is the SVM hinge loss, C and λ are regularization parameters, $\Omega(\cdot)$ is a differentiable regularization function of $\mathbf{F}$ and $g(\cdot)$ is the decision function defined here for the filtered pixel $\mathbf{p}$:

$$g(\widetilde{\mathbf{X}}_{\mathbf{p},.}) = \sum_{\mathbf{q}\in\mathcal{S}_l} \alpha_\mathbf{q} \mathbf{Y}_\mathbf{q} k(\widetilde{\mathbf{X}}_{\mathbf{q},.}, \widetilde{\mathbf{X}}_{\mathbf{p},.}) + b, \quad (4)$$

where $\alpha_\mathbf{p}$ are the dual variables of problem (3), $k(\cdot, \cdot)$ is a valid kernel function as the RBF function defined above and b is the bias term.

Note that the two leftmost terms in (3) correspond to a SVM problem for filtered pixels in $\widetilde{\mathbf{X}}$. Thus this problem implicitly depends on the variable $\mathbf{F}$. This objective function is non-convex with respect to both variables. However, for a fixed $\mathbf{F}$, the optimization problem with respect to $g(\cdot)$ is convex and reduces to a SVM problem on the filtered image. Let us denote the objective function of Eq. (3) as $J(\mathbf{F})$. Then this problem reduces to minimize with regard to $\mathbf{F}$:

$$J = J'(\mathbf{F}) + \lambda\Omega(\mathbf{F}) \quad (5)$$

with:

$$J'(\mathbf{F}) = \min_g \left\{ \frac{1}{2}\|g\|^2 + \frac{C}{n} \sum_{\mathbf{p}\in\mathcal{S}_l} H(\mathbf{Y}_\mathbf{p}, g(\widetilde{\mathbf{X}}_{\mathbf{p},.})) \right\} \quad (6)$$

where $g(\cdot)$ is defined in Equation (4). Due to the strong duality of the SVM problem, $J'(\cdot)$ can be expressed in dual form as the constrained maximization problem:

$$\max_\alpha \left\{ -\sum_{(\mathbf{p},\mathbf{q})\in\mathcal{S}_l^2} \mathbf{Y}_\mathbf{p}\mathbf{Y}_\mathbf{q}\alpha_\mathbf{p}\alpha_\mathbf{q}\widetilde{K}_{\mathbf{p},\mathbf{q}} + \sum_\mathbf{p} \alpha_\mathbf{p} \right\} \quad (7)$$

$$\text{s.t.} \quad \frac{C}{n} \geq \alpha_\mathbf{p} \geq 0 \quad \forall \mathbf{p} \in \mathcal{S}_l \quad \text{and} \quad \sum_\mathbf{p} \alpha_\mathbf{p}\mathbf{Y}_\mathbf{p} = 0$$

with α the dual variables used in the decision function (4). According to [20], and for a given $\mathbf{F}^*$, $J'(\cdot)$ is differentiable with respect to $\mathbf{F}$. This means that at the point $\mathbf{F}^*$, the gradient of $J(\cdot)$ can be computed.

In [19], the authors proposed to solve the problem in (5) with a gradient descent on $J(F)$ along $\mathbf{F}$. In the non-linear case, and when an RBF kernel is used, the gradient of $J'(\cdot)$ with respect to $\mathbf{F}$ is:

$$\begin{aligned}\nabla J(\mathbf{F})_{u,v,k} &= \textstyle\sum_{(\mathbf{p},\mathbf{q})}(\mathbf{X}_{\mathbf{p}+(u,v)-f_0,k} - \mathbf{X}_{\mathbf{q}+(u,v)-f_0,k}) \\ &\times(\widetilde{\mathbf{X}}_{\mathbf{p},k} - \widetilde{\mathbf{X}}_{\mathbf{q},k})\widetilde{K}_{\mathbf{p},\mathbf{q}}\mathbf{Y}_{\mathbf{p}}\mathbf{Y}_{\mathbf{q}}\alpha^*_{\mathbf{p}}\alpha^*_{\mathbf{q}},\end{aligned}$$

where α^* is the SVM solution for a fixed $\mathbf{F}$. Note that this sum needs to be done only for the pixel $\mathbf{p}$ that are support vectors.

2.3. Regularization and Algorithm

The most important part of our method is the choice of the regularization term $\Omega(\cdot)$ in (5), that must be easily differentiable in order to compute the gradient of $J(F)$. Similarly to [19], we use here a regularization term based on the Frobenius norm of the 3-dimensional matrix $\mathbf{F}$:

$$\Omega(\mathbf{F}) = \sum_{u,v,k}^{f,f,d} \mathbf{F}^2_{u,v,k} \tag{8}$$

This regularization term is differentiable and the gradient is easy to compute. Minimizing this regularization term can be seen as minimizing the filter's energy. In terms of classification, the filter matrix can be seen as a kernel parameter weighting neighboring pixels. Moreover, the Gaussian kernel defined in (2) shows that the component-wise convolution can be seen as a scaling of the spectral channels before kernel computation. For this reason, the kernel parameter σ does not need to be optimized in the procedure and can be set to the value of 1.

The intuition of how this regularization term influences the filter learning is the following: suppose that the decision function $g(\cdot)$ has been learned by minimizing only $J'(\cdot)$. In this case, the learned filter matrix will simply maximize the margin between classes. But adding the Frobenius regularizer will force non-discriminative filter coefficients to vanish, thus minimizing their impact on the kernel of the neighboring pixels.

In order to solve the optimization problem, we propose a conjugate gradient (CG) descent algorithm along $\mathbf{F}$ with a line search for finding the optimal step. The method is detailed in Algorithm 1, where β is the CG update parameter and $D^i_{\mathbf{F}}$ represents the descent direction for the ith iteration. For the experiments, we used the β proposed by Fletcher and Reeves (see [21]). In the implementation proposed, the iterations in the algorithm may be stopped by two criteria:

Algorithm 1 KF-SVM solver

Set $\mathbf{F}_{u,v} = 1/f$ for $v = 1 \cdots d$ and $u = 1 \cdots f$
Set i=0, Set $D^0_F = 0$
repeat
  i=i+1
  $G^i_F \leftarrow$ gradient of $J'(\mathbf{F}) + \lambda\Omega(\mathbf{F})$ wrt. $\mathbf{F}$
  $\beta \leftarrow \frac{\|G^i_F\|^2}{\|G^{i-1}_F\|^2}$ (Fletcher and Reeves)
  $D^i_F \leftarrow -G^i_F + \beta D^{i-1}_F$
  $(\mathbf{F}^i, \alpha*) \leftarrow$ Line-Search along D^i_F
until Stopping criterion is reached

either a threshold on the relative variation of $J(F)$ or on the norm of the variation of the coefficients in $\mathbf{F}$.

Due to the non-convexity of the objective function, it is difficult to provide an exact evaluation of the solution complexity. However, we know that the gradient computation is $\mathcal{O}(n_s^2 \cdot f^2)$ with n_s being the number of support vectors. When $J(F)$ is computed in the line search, a $\mathcal{O}(n^2 \cdot d)$ SVM is solved and a $\mathcal{O}(n \cdot f \cdot d)$ filtering is applied. Note that, in order to speed up the method, warm-start initialization is used when using iteratively the SVM solver.

3. DATASETS AND EXPERIMENTAL SETUP

The proposed method is tested on a VHR QuickBird image of the city of Zurich, Switzerland (see Fig. 1). The dataset considered represents a residential area in the South-West part of the city. Seven classes were labeled by photointerpretation. The main challenge is to distinguish between the two classes of buildings and of roads by applying spatial filtering, because the spectral difference between these couples of classes is really low.

In the experiments, four models have been compared: i) a standard SVM, ii) a SVM that considers pixels filtered by local averaging with a $f \times f$ moving window (AvgSVM

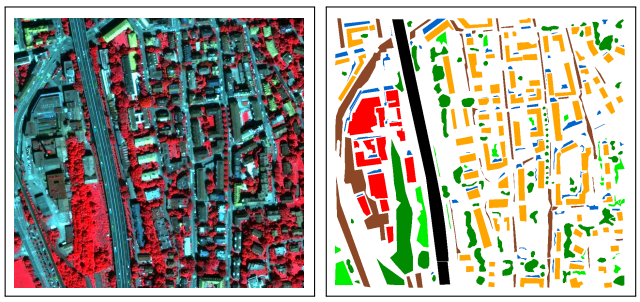


Fig. 1. QuickBird scene of suburbs of Zurich (left) and labeled pixels (right). Legend: *dark green = trees; light green = meadows; black = speedway; brown = roads; orange = residential buildings; red = commercial buildings; blue = shadows.*

Table 1. Results for binary classification.

Method	Class	Training Pixels	#Class Pixels	AUC	Kappa
SVM	Residential	~ 5000	2000	0.904	0.638
AvgSVM				0.916	0.689
WinSVM				**0.947**	0.730
KF-SVM				0.938	**0.742**
SVM	Buildings*	~ 4000	1000	0.938	0.706
AvgSVM				0.946	0.779
WinSVM				0.970	0.807
KF-SVM				**0.974**	**0.815**

* Pixels from classes 'Residential' and 'Commercial'.

Table 2. Results for multiclass classification.

Method	Classes	Filter size	Training Pixels	[%]OA	Kappa
SVM	7	9	~ 5000	75.11	0.685
AvgSVM				83.68	0.796
WinSVM				82.98	0.785
KF-SVM				**85.32**	**0.816**
SVM	6*	9	~ 5000	83.04	0.772
AvgSVM				89.48	0.860
WinSVM				**91.71**	**0.889**
KF-SVM				91.45	0.885

* Pixels from classes 'Residential' and 'Commercial'.

hereafter), iii) a SVM using a stacked vector of the reflectance values of the local $f \times f$ neighborhood (WinSVM [19] hereafter) and iv) the proposed KF-SVM, learning a filter of the same size for each channel and classifier. From the labeled pixels available in the ground survey, $4,000$ to $5,000$ are used for training, $5,000$ for validation, and $26,000$ for testing. RBF kernels have been used in all cases. Regarding the kernel parameter σ, it has been optimized by cross-validation for the SVM and AvgSVM methods, while for KF-SVM it is kept to a standard value of 1, since its optimal value is learned directly from the filter: as each band of the image is filtered before the computation of the kernels, it is always possible to find a regularization parameter λ returning a solution which is equivalent to the one that would have been obtained when using an optimal kernel value (see Section 2.3 for more details). In conclusion, C and σ parameters have been validated for the SVM, AvgSVM and WinSVM methods while C and λ have been validated for KF-SVM.

4. RESULTS AND DISCUSSION

In this Section, the experiments are presented for two classification scenarios: firstly, a binary classification problem is considered. Secondly, a multiclass setting, where as many one-against-all binary SVM as the number of classes are solved.

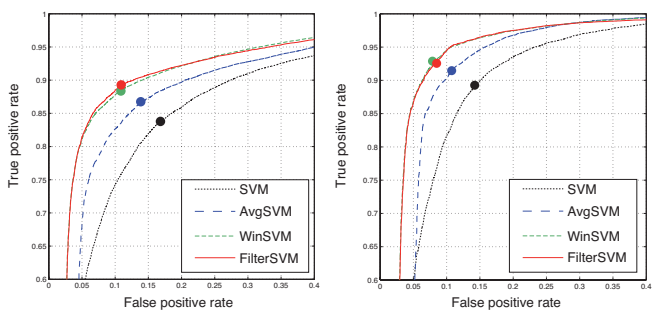
4.1. Binary classification

Table 1 shows the results for the binary problem with classes 'Residential buildings' and 'Buildings'. For the latter, we merged the residential and commercial buildings classes. The numerical results show good performance of KF-SVM, that strongly improves the AvgSVM result, with an increase between 4 and 6% in the estimated kappa statistic. With respect to WinSVM, the proposed method provides similar results, with a slight improvement of the Kappa coefficient, confirming the observations done for the linear case in [19].

Regarding the ROC curve, the AUC is only slightly improved by the proposed approach, since the detection is already very good with the mean filter. Nonetheless, by looking at the curves of Fig. 2, KF-SVM results in a global improvement of the results of SVM and AvgSVM, which can be characterized by an increase in the location of the γ–point, i.e. the closest point to the (0,1) corner of the ROC. In this sense, KF-SVM maximizes the true positive rate (tpr) while minimizing the false positive rate (fpr). Compared to WinSVM, the γ–points of the two methods are practically identical for both classification problems.

4.2. Multiclass classification

Multiclass classification results are shown in Table 2: for the 7–classes setting, the inclusion of spatial information strongly improves the results of the SVM, whose overall accuracy is increased by +8–10%. This result was expected, because the use of contextual information allows to discriminate the classes with no strong spectral differences, cf. Section 3. This can be seen in the classification maps reported in Fig. 3: KF-SVM allows the detection of the class 'Com-



(a) *(b)*

Fig. 2. Details of the ROC curves for the (a) 'Residential Buildings' and (b) 'Buildings' targets. The dots correspond to the γ-points, $\gamma = \arg\min\{fpr^2 + (1 - tpr)^2\}$.

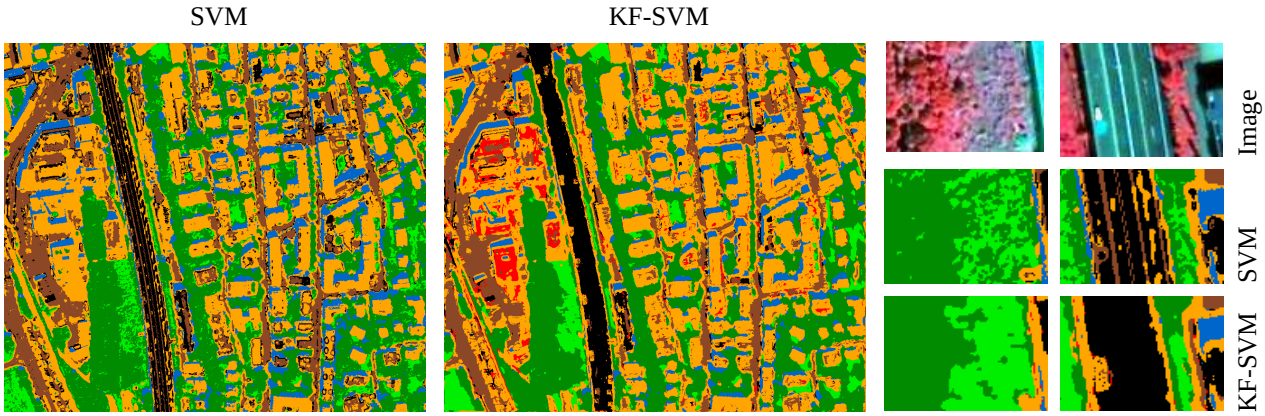


Fig. 3. Classification maps for the SVM and the proposed KF-SVM, for the whole scene and two interesting regions.

mercial buildings', whose difference with respect to the residential buildings in uniquely related to the size of the objects. Moreover, KF-SVM enforces a smooth classification result and avoids noisy classification maps. With respect to the mean filter, KF-SVM provides a gain of about +2% in both accuracy and estimated kappa statistic, suggesting that an appropriate filter has been learned from data, at the same time avoiding the oversmoothing of the mean filter and taking advantage of the spatial context of the pixel. The same order of gain is observed with respect to WinSVM.

Regarding the 6–classes setting, the standard SVM model leads to better results, since the size variability of buildings is no more taken into account. Nonetheless, both the filter approaches still improve the SVM result by +6–8% and the difference of about 2% between the two filters holds. For this easier 6 classes setting, WinSVM (which is convex) performs optimally and provides the best results, that slightly outperform the proposed KF-SVM (+ 0.5% in overall accuracy), which only proposes an approximated solution.

4.3. On filter weights

Unlike blackbox approaches as SVM or WinSVM, the proposed KF-SVM provides a smoothing filter maximizing margin separation. Such discriminative filters can be further used, for instance to train a second classifier using them as features. This opens a set of challenging opportunities to exploit the convolution obtained. A first step is to analyze the filter's weights, as they are learned directly from the data. Training the classifier returns a proper convolution filter whose weights represent the signal characteristics. A possible statistic summarizing the filter is the Frobenius norm of the filter *per* spectral channel, as discussed in Section 2.3. Figure 4 shows the variation of the Frobenius norm coefficients, in (8), during the optimization of the model for the classification of the particularly interesting 'Buildings' class. Note that, for this problem, the near infrared band is enhanced in the mixture, since it carries the most suitable information to discriminate the buildings from the vegetation classes.

5. CONCLUSION

In this paper, we presented an algorithm learning the appropriate convolutional filter for the classification of VHR remote sensing images. The algorithm jointly learns a SVM classifier and a 2D filtering of the image in order to maximize the margin between the classes. The algorithm has shown good results in two challenging remote sensing problems: VHR imagery multiclass classification and detection in urban settlements. KF-SVM showed similar results with respect to a contextual SVM using a stacked features approach (WinSVM). However, these results are obtained using much lower dimensional vectors and hence the mehtod is more robust to overfitting. Also, this may be of great importance if the filter is learned as a preprocessing step before using a Bayesian or GMM classifier: in this case a method using the stacked approach might fail because of the higher

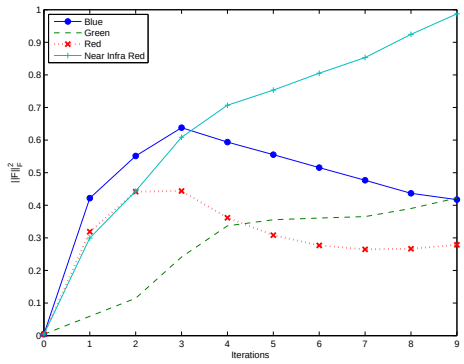


Fig. 4. Variation of the per-channel Frobenius norm during the KF-SVM optimization.

dimensionality induced [22].

Here we included the KF-SVM in one-versus-all schemes but, the optimization in one-versus-one schemes would report filters optimized to discern among pairs of classes. Such filters could help solving problems with highly overlapping classes in conjunction with standard SVM. Besides, a possible improvement of the current algorithm should take into account the local characteristics of the image, returning pointwise convolution filters. These will be studied in the future since they certainly are issues for most aerial and VHR satellite image.

Acknowledgment

The authors would like to thank Prof. M. Kanevski (IGAR, University of Lausanne) for providing the QuickBird data.

6. REFERENCES

[1] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *5th ACM Workshop on Computational Learning Theory*, Pittsburgh, USA, 1992, pp. 144–152.

[2] V. Vapnik, *Statistical Learning Theory*, Wiley-Interscience, NJ, USA, 1998.

[3] G. Camps-Valls and L. Bruzzone, *Kernel methods in Remote Sensing Image Processing*, Wiley and Sons, 2009.

[4] P. Soille, *Morphological Image Analysis: Principles and Applications*, Germany:Springer-Verlag, 2003.

[5] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, Sept. 2003.

[6] J. Inglada, "Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features," *ISPRS Journal of Photogrammetry Rem. Sens.*, vol. 62, pp. 236–248, 2007.

[7] R. C. Dubes and A. K. Jain, "Random field models in image analysis," *J. Appl. Stat.*, vol. 16, no. 2, pp. 131–163, 1989.

[8] Q. Jackson and D. Landgrebe, "Adaptive bayesian contextual classification based on Markov random fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 11, pp. 2454–2463, Mar. 2002.

[9] I. Tomas, "Spatial postprocessing of spectrally classified Landsat data," *Photogrammetic Engineering and Remote Sensing*, vol. 46, pp. 1201–1206, 1980.

[10] J. Ton, J. Sticklen, and A. Jain, "Knowledge-based segmentation of Landsat images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 29, pp. 222–231, Mar 1991.

[11] B. Solaiman, R. Koffi, M. Mouchot, and A. Hillion, "An information fusion method for multispectral image classification postprocessing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 2, pp. 395–406, Mar 1998.

[12] Y. Zhang, "Detection of urban housing development by fusing multisensor satellite data and performing spatial feature post-classification," *International Journal of Remote Sensing*, vol. 22, no. 17, pp. 3339–3355, 2001.

[13] G. Camps-Valls, L. Gómez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, Jan 2006.

[14] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simple MKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[15] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple kernel classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, 2010.

[16] D.A. Clausi and B. Yue, "Comparing cooccurrence probabilities and Markov random fields for texture analysis of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 1, pp. 215–228, Jan. 2004.

[17] 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.

[18] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multi-source composite kernels for urban image classification," *IEEE Geosci. Remote Sens. Lett., Special Issue ESA_EUSC*, vol. 7, no. 1, pp. 88–92, 2010.

[19] R. Flamary, B. Labbé, and A. Rakotomamonjy, "Large margin filtering for signal sequence labeling," in *IEEE Conference on Acoustic, Speech and Signal Processing ICASSP*, Austin, Texas, USA, 2010.

[20] J.F. Bonnans and A. Shapiro, "Optimization problems with pertubation : A guided tour," *SIAM Review*, vol. 40, no. 2, pp. 202–227, 1998.

[21] W.W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific journal of Optimization*, vol. 2, no. 1, pp. 35–58, 2006.

[22] R. Flamary, B. Labbé, and A. Rakotomamonjy, "Filtrage vaste marge pour l'étiquetage séquentiel de signaux," in *Conference en Apprentissage CAp*, 2010.