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, 2010. <hal-00528923>
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 cooccurrence 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
a particular gradient-descent approach [19]. Since the pre-
processing phase, i.e. the definition and selection of the fil-
ters, is integrated into the learning process, the proposed ap-
proach 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 sce-
narios: 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 dis-
cussed 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 con-
volution of the signals to enhance signal sequence labeling.
In this paper, this approach is extended to pixel classi-
fication of remote sensing multidimensional images: in this
case, the filtering corresponds to learning a convolution ker-
nel that promotes maximum margin between classes. Ad-
ditionally, the the filtered SVM proposed in [19] is here ex-
tended to the non-linear case, through Gaussian kernels.

2.1. Definitions

Consider an image $X \in \mathbb{R}^{r_1 \times r_2 \times d}$, 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 nota-
tion, we define $X_{i,j,k} = X_{p,k}$ as the $k$th component of pixel
$p = (i,j)$. Since the approach proposed is supervised, a list
of $n$ labeled pixels $p = (i,j) \in S_l$ is available, as well as a
corresponding labels vector $Y_p \in \{-1,1\}, \forall p \in S_l$.

As stated in the introduction, a 2-dimensional convolu-
tion may be applied to the image in order to enhance dis-
criminative information. In this case, the convolution is
performed by applying per component convolution matrices
that are stored in $F \in \mathbb{R}^{r_1 \times r_2 \times d}$, where $f$ is the size of
the bidimensional square filters. When convoluting, the fil-
ter is centered on the current pixel. The filtered image $X$ is
defined by:

$$\tilde{X}_{p,k} = \sum_{u=1}^{f,f} F_{u,v,k} X_{i+u-f_o,j+v-f_o,k}$$

$$= \sum_{u=1}^{f,f} F_{u,v,k} X_{(u,v)-f_o,k}$$

(1)

where $f_0 = f/2$. The filter coefficients in $F$ define the type
of smoothing/enhancement performed by the convolution.

For instance, smoothing can be achieved by applying an av-
erage filter corresponding to a matrix $F_{i,j,k} = \frac{1}{r_2}, \forall (i,j,k)$.

Finally, we define the RBF kernel between the filtered
pixels $p$ and $q$:

$$K_{p,q} = \exp \left( -\frac{||\tilde{X}_{p,k} - \tilde{X}_{q,k}||^2}{2\sigma^2} \right)$$

(2)

where $\sigma$ 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 con-
textual 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 con-
straints. This double objective can be formulated with the
following minimization problem:

$$\min_{g,F} \left\{ \frac{1}{2} ||g||^2 + \frac{C}{n} \sum_{p \in S_l} H(Y_p, g(\tilde{X}_{p,k})) + \lambda \Omega(F) \right\}$$

(3)

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

$$g(\tilde{X}_{p,k}) = \sum_{q \in S_l} \alpha_q Y_q k(\tilde{X}_{q,k}, \tilde{X}_{p,k}) + b,$$

(4)

where $\alpha_p$ 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 $\tilde{X}$. Thus this problem
implicitly depends on the variable $F$. This objective func-
tion is non-convex with respect to both variables. However,
for a fixed $F$, 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 $J(F)$. Then
this problem reduces to minimize with regard to $F$:

$$J = J'(F) + \lambda \Omega(F)$$

(5)

with:

$$J'(F) = \min_g \left\{ \frac{1}{2} ||g||^2 + \frac{C}{n} \sum_{p \in S_l} H(Y_p, g(\tilde{X}_{p,k})) \right\}$$

(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:

$$\max_{\alpha} \left\{ - \sum_{(p,q) \in S_l^2} Y_p Y_q \alpha_p \alpha_q K_{p,q} + \sum_{p} \alpha_p \right\}$$

s.t. $\frac{C}{n} \geq \alpha_p \geq 0 \quad \forall p \in S_l$ and $\sum \alpha_p Y_p = 0$
with $\alpha$ the dual variables used in the decision function (4). According to [20], and for a given $F^*$, $J'(\cdot)$ is differentiable with respect to $F$. This means that at the point $F^*$, 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 $F$. In the non-linear case, and when an RBF kernel is used, the gradient of $J'(\cdot)$ with respect to $F$ is:

$$
\nabla J(F)_{u,v,k} = \sum_{(p,q)}(X_{p+(u,v)}-f_{a,k} - X_{q+(u,v)}-f_{a,k})
\times(\tilde{X}_{p,k} - \tilde{X}_{q,k})\tilde{K}_{p,q}Y_p Y_q \alpha^*_p \alpha^*_q,
$$

where $\alpha^*$ is the SVM solution for a fixed $F$. Note that this sum needs to be done only for the pixel $p$ 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 $F$:

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

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 $\sigma$ 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 $F$ with a line search for finding the optimal step. The method is detailed in Algorithm 1, where $\beta$ is the CG update parameter and $D_F^i$ represents the descent direction for the $i$th iteration. For the experiments, we used the $\beta$ 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 $F_{u,v} = 1/f$ for $u = 1 \cdots d$ and $a = 1 \cdots f$

Set $i=0$, Set $D_F^i = 0$

repeat

i=i+1

$G_F^i \leftarrow$ gradient of $J'(F) + \lambda \Omega(F)$ wrt. $F$

$\beta \leftarrow \frac{\|G_F^i\|}{\|G_F^i\|^2}$ (Fletcher and Reeves)

$D_F^i = -G_F^i + \beta D_F^{i-1}$

$(F^i, \alpha^*) \leftarrow$ Line-Search along $D_F^i$

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 $F$.

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 $O(n^2 \cdot f^2)$ with $n_s$ being the number of support vectors. When $J(F)$ is computed in the line search, a $O(n^2 \cdot d)$ SVM is solved and a $O(n \cdot f \cdot d)$ 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 couple 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.
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-

**Table 1. Results for binary classification.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Class</th>
<th>Training Pixels</th>
<th>#Class Pixels</th>
<th>AUC</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>Residential</td>
<td>~ 5000</td>
<td>2000</td>
<td>0.904</td>
<td>0.638</td>
</tr>
<tr>
<td>AvgSVM</td>
<td>Residential</td>
<td>~ 5000</td>
<td>2000</td>
<td>0.916</td>
<td>0.689</td>
</tr>
<tr>
<td>WinSVM</td>
<td>Residential</td>
<td>~ 5000</td>
<td>2000</td>
<td>0.947</td>
<td>0.730</td>
</tr>
<tr>
<td>KF-SVM</td>
<td>Residential</td>
<td>~ 5000</td>
<td>2000</td>
<td>0.938</td>
<td>0.742</td>
</tr>
<tr>
<td>SVM</td>
<td>Buildings</td>
<td>~ 4000</td>
<td>1000</td>
<td>0.938</td>
<td>0.706</td>
</tr>
<tr>
<td>AvgSVM</td>
<td>Buildings</td>
<td>~ 4000</td>
<td>1000</td>
<td>0.946</td>
<td>0.779</td>
</tr>
<tr>
<td>WinSVM</td>
<td>Buildings</td>
<td>~ 4000</td>
<td>1000</td>
<td>0.970</td>
<td>0.807</td>
</tr>
<tr>
<td>KF-SVM</td>
<td>Buildings</td>
<td>~ 4000</td>
<td>1000</td>
<td>0.974</td>
<td>0.815</td>
</tr>
</tbody>
</table>

* Pixels from classes ‘Residential’ and ‘Commercial’.

**Table 2. Results for multiclass classification.**

<table>
<thead>
<tr>
<th>Method</th>
<th>Classes</th>
<th>Filter size</th>
<th>Training Pixels</th>
<th>[%]OA</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>7</td>
<td>9</td>
<td>~ 5000</td>
<td>75.11</td>
<td>0.685</td>
</tr>
<tr>
<td>AvgSVM</td>
<td>7</td>
<td>9</td>
<td>~ 5000</td>
<td>83.68</td>
<td>0.796</td>
</tr>
<tr>
<td>WinSVM</td>
<td>7</td>
<td>9</td>
<td>~ 5000</td>
<td>82.98</td>
<td>0.785</td>
</tr>
<tr>
<td>KF-SVM</td>
<td>7</td>
<td>9</td>
<td>~ 5000</td>
<td>85.32</td>
<td>0.816</td>
</tr>
<tr>
<td>SVM</td>
<td>6</td>
<td>9</td>
<td>~ 5000</td>
<td>83.04</td>
<td>0.772</td>
</tr>
<tr>
<td>AvgSVM</td>
<td>6</td>
<td>9</td>
<td>~ 5000</td>
<td>89.48</td>
<td>0.860</td>
</tr>
<tr>
<td>WinSVM</td>
<td>6</td>
<td>9</td>
<td>~ 5000</td>
<td>91.71</td>
<td>0.889</td>
</tr>
<tr>
<td>KF-SVM</td>
<td>6</td>
<td>9</td>
<td>~ 5000</td>
<td>91.45</td>
<td>0.885</td>
</tr>
</tbody>
</table>

* Pixels from classes ‘Residential’ and ‘Commercial’.

![Fig. 2. Details of the ROC curves for (a) ‘Residential Buildings’ and (b) ‘Buildings’ targets. The dots correspond to the γ-points, γ = arg min{fpr² + (1 - tpr)²}.](image-url)
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 method 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

![Figure 3](image1.png)

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

![Figure 4](image2.png)

**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


