Semisupervised Classification of Remote Sensing Images with Hierarchical Spatial Similarity

Lian-Zhi Huo, *Member, IEEE*, Ping Tang, Zheng Zhang and Devis Tuia, *Member, IEEE*

Abstract

A semisupervised kernel deformation function including spatial similarity is proposed for the classification of remote sensing images. The method exploits the characteristic of these images, in which spatially nearby points are likely to belong to the same class. To fulfill this assumption, a kernel encoding both spatial and spectral proximity using unlabeled samples is proposed. In this letter, two similarity functions for constructing a spatial kernel are proposed. Experimental tests are performed on very high resolution multispectral and hyperspectral data. With respect to state-of-the-art semisupervised methods for remote sensing images, the proposed method incorporating spatial similarity obtains higher classification accuracies and smoother classification maps.

Index Terms

Image segmentation, image classification, kernel methods, support vector machines (SVM).

Manuscript received November 7, 2013; revised March 9,2014 and May 14; and accepted June 3, 2014.

This work has been partly supported by the Youth Fund Program (grant Y3SJ7700CX) and the ”135” Strategy Planning (grant Y3SG1500CX) of RADI, CAS and partly by the Swiss National Science Foundation (grant PZ00P2-136827).

Lian-Zhi Huo, Ping Tang and Zheng Zhang are with the Institute of Remote Sensing and Digital Earth (RADI), Chinese Academy of Sciences, Chaoyang District, Beijing, 100101 China. E-mail: huolz@radi.ac.cn; tangping@radi.ac.cn; zhangzheng2035@163.com.

Devis Tuia is with the Laboratory of Geographic Information Systems, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland. E-mail: devis.tuia@epfl.ch.

I. Introduction

Remote sensing (RS) plays an important role in the Earth sciences by providing images covering wide areas with fast revisiting and in a non-intrusive way. To make the images usable for monitoring purposes, the images are often converted to land-cover maps. To do so, supervised classification methods are mainly utilized, due to their more accurate classification results. However, the success of supervised classification methods depends greatly on the availability and quantity of labeled samples. Without enough labeled samples, the underlying probability distribution function of the image cannot be properly described [1]. One way to improve the quality of modeling is to consider the information contained in the numerous unlabeled samples and use it in conjunction to the one carried by the labeled pixels. These methods are called semisupervised learning (SSL) methods.

SSL is a very active field of machine learning and has attracted wide research interests [2]. In RS, SSL applies naturally, principally due to the availability of unlabeled samples in the images and to the smooth manifolds where the images live [3]. Among the most successful SSL methods proposed in RS, Transductive Support Vector Machines (TSVM) [4] use several semilabeled pixels selected based on the current classification hyperplanes and maximize the margin for both labeled and semilabeled samples. Other SSL methods encode manifold consistency by using graphs: in [5], the information contained in the labeled pixels is propagated to their neighbors until the whole image attains a global steady state; in [6], the SVM decision function is modified by the local consistency encoded by the graph Laplacian, thus forcing neighbors in the feature space to have similar decision function values.

Later studies have proposed to deform the kernel prior to SVM optimization. This seems a convenient way to encode the information carried by unlabeled samples, since no modification of the classifier must be performed, as the information of the unlabeled samples is already included in the kernel matrix. In [7], the authors propose to build an unsupervised kernel, a *Bagged kernel*, using clustering, where similarity is assessed by the co-occurrence of cluster membership of the samples considered. Tuia *et al.* reformulated this idea for image classification [1] and later proposed a multiscale version of the Bagged kernel [8]. In all these studies, the kernel based on unlabeled samples is built and then combined with the standard SVM kernel by kernel composition [9], [10].

Despite their good performances and simplicity of application, these methods seldom consider spatial characteristics of RS images. Since they represent objects in the geographical space, RS images are spatially continuous and the objects observed tend to be spatially smooth. If one pixel belongs to one land cover type, e.g., cropland, it is plausible that the spatially nearby pixels also belong to cropland. Similarly to the cluster assumption in SSL (pixels of the same class will belong to the same cluster in the feature space [2]), we can refer to this assumption as spatial cluster assumption. In [11], the authors included spatial information in a progressive SSL classifier, but limited the experimentation to a single type and scale of spatial filter, while RS objects at VHR are typically multiscale. In [12], the authors proposed the selection of unlabeled samples based on spatial proximity.

Another possibility is to include spatial smoothness via image segmentation. In [13], Li *et al.* modeled spatial information with Markov Random Field in semisupervised segmentation method. In [14], authors enrich the feature vector with features extracted by hierarchical segmentation and then train a SVM in the enriched space. In [15], interactive learning is used to find the best segmentation exploiting a limited number of labeled pixels and a hierarchical segmentation result.

In this paper we use hierarchical segmentation to define a new measure of similarity between pixels and improve the quality of classification with kernel classifiers: we propose a semisupervised SVM kernel encoding the spatial consistency assumption. Spatial consistency is achieved by a similarity measure issued from a hierarchical segmentation of the image, which provides a hierarchical description of the data. Two distance measures for constructing the spatial kernel are proposed: the first hierarchical level in which two pixels belong to the same cluster and a similarity based on the aggregation distance at that level of segmentation. The final kernel is obtained by linearly combing the spatial kernel with the Bagged kernel as in [1]. The proposed method provides *out-of-sample predictions* and has a training complexity equivalent to the traditional SVM, plus the segmentation cost.

The main contributions of this letter are 1) the use of hierarchical segmentation to model spatial consistency of RS images in SSL methods; and 2) the introduction of two distance measures based on hierarchical segmentation to define the spatial kernel. In what follows, Section II presents the proposed SSL method. Section III presents the datasets used, as well as the experimental setup. Section IV reports and analyzes experimental results. Section V concludes the paper.

II. The Proposed Algorithm

The proposed algorithm is a standard SVM with a deformed kernel. The deformation is justified by the need of designing a kernel function, which represents the similarity among samples in an accurate way [1]: following the cluster assumption in SSL [2], we want to increase the similarity between pixels belonging to the same cluster, in both spectral and spatial terms. Regarding spectral consistency, a way to enforce the cluster assumption is to design a Bagged kernel (or cluster kernel [7]), where the traditional RBF kernel, K_{RBF}, is deformed by a second kernel encoding the cluster relations among unlabeled samples, K_{BAG}, as: $K(\mathbf{x}_i, \mathbf{x}_j) = \alpha K_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j) + (1-\alpha)K_{\text{BAG}}(\mathbf{x}_i, \mathbf{x}_j)$, with $0 < \alpha < 1$.

K_{BAG} is the Bagged kernel constructed from unlabeled data:

$$K_{\text{BAG}}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{T}\sum_{t=1}^{T}[c_i^t = c_j^t] \tag{1}$$

where T is the number of realizations of a clustering algorithm, c_i^t is the cluster membership of pixel $\mathbf{x}_i$ at realization t and operator $[c_i^t = c_j^t]$ returns "1" if sample $\mathbf{x}_i$ and $\mathbf{x}_j$ belong to the same cluster at realization t and "0" otherwise. The kernel resulting from Eq. (1) is a valid kernel [7].

The kernel proposed in this paper extends this formulation, by adding a third term increasing similarity for spatially smooth regions obtained with hierarchical segmentation. After running a hierarchical segmentation algorithm on the entire image (thus using all unlabeled and labeled samples available), a new kernel encoding spatial consistency, K_{SEG}, is built and linearly combined to the previous K_{RBF} and K_{BAG}. The final kernel K is obtained as follows:

$$\begin{gathered} K = \alpha K_{\text{RBF}} + \beta K_{\text{BAG}} + (1-\alpha-\beta)K_{\text{SEG}} \\ (0 < \alpha < 1, 0 < \beta < 1, 0 \leq \alpha + \beta \leq 1) \end{gathered} \tag{2}$$

The algorithm's computational cost is the sum of i) the hierarchical segmentation algorithm (run once), ii) the cost of the clustering algorithm (in our case T runs of k-means, $\mathcal{O}(mdk)$) and iii) the standard SVM cost ($\mathcal{O}(n^2 d)$). Algorithm 1 summarizes the proposed method. The next section details the kernel encoding spatial consistency, K_{SEG}.

A. Spatial Kernel Construction

Let s_i^h be the segment to which the pixel $\mathbf{x}_i$ belongs at hierarchical level h. The hierarchy has been obtained running a segmentation algorithm on the entire image and $h = 1$ denotes

Algorithm 1 Bagged SVM with hierarchical spatial similarity

Inputs

- Labeled pixels $X^L = \{\mathbf{x}_i, y_i\}_{i=1}^n \in \mathbb{R}^d$;
- Unlabeled pixels $X^U = \{\mathbf{x}_j\}_{j=1}^m \in \mathbb{R}^d$, $m >> n$;
- Number of realizations of the clustering algorithm, T.

1: run a hierarchical segmentation algorithm on the entire image. Store the hierarchy;
2: compute K_{SEG} for pixels in X^L according to Eq. (3);
3: **for** $t = 1, \ldots, T$ **do**
4: Run a clustering algorithm on X^U. Store centers $\mathbf{x}_c^{(t)}$;
5: Assign cluster membership c to samples in X^L;
6: **end for**
7: Compute K_{BAG} according to Eq. (1);
8: Compute the final kernel K with Eq. (2);
9: Run standard SVM using the deformed kernel;

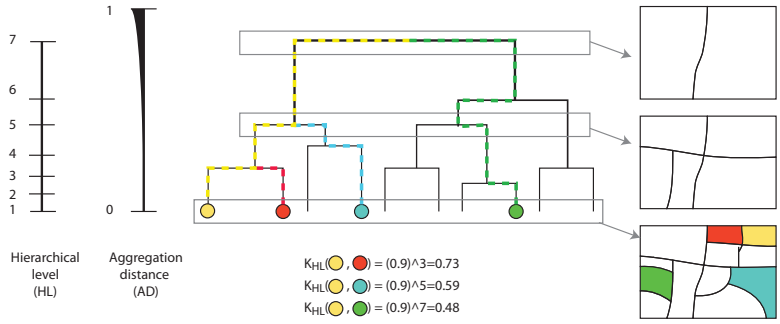


Fig. 1. Illustration of spatial kernel construction with the hierarchical description of the image obtained by segmentation. The yellow node and red nodes are first merged as one segment in the 3rd hierarchical level. So, according to the HL distance and Eq. (3) with $a = 1$, the similarity between these two samples is 0.73, which is higher than for the pixels belonging to other objects, as the cyan and green ones.

the finest segmentation level in the hierarchical sequence. For a fixed number of hierarchical levels H, each pixel $\mathbf{x}_i$ can be represented using a sequence of region labels $\mathbf{s}_i = \{s_i^h\}_{h=1}^H$. A spatial measure of similarity can be defined using this sequence: for two pixels $\mathbf{x}_i$ and $\mathbf{x}_j$, spatial similarity can be measured by searching for the first common segment they belong to. If such

segment is low in the hierarchy, the pixels are considered to be similar. An illustration of the spatial kernel construction is shown in Fig. 1.

We propose two distance measures to assess spatial similarity between pixels: the hierarchical level (HL) and the aggregation distance for a given level of aggregation (AD). The HL distance is defined as $d_{\text{HL}}(\mathbf{x}_i, \mathbf{x}_j) = \text{argmin}_{\text{h}}\{s_i^h = s_j^h\}$. In other words, it corresponds to the smallest level of aggregation, where the pixels belong to the same segment: if two pixels belong to the same object at the smallest level ($s_i^1 = s_j^1$), we will observe $d_{\text{HL}}(\mathbf{x}_i, \mathbf{x}_j) = 1$. The AD is defined as the aggregation distance corresponding to the hierarchical level h (i.e., the distance value at which two segments are merged in the hierarchical level h): in the case above $d_{\text{AD}}(\mathbf{x}_i, \mathbf{x}_j) = 0$. These two distance measures have a different range (e.g., if we have 300 levels of hierarchy, HL is from 1 to 300, while AD is from 0 to 1 in the case of spectral angle mapper distance) and different meanings, as HL is linear with respect to the aggregation (each aggregation produces the same increase in HL, as in Fig. 1), while AD respects the jumps in similarity between the segments (first merging are usually done at small AD, while the last ones merge dissimilar segments). In both cases, the final spatial kernel K_{SEG} is defined as a monotonically decreasing function versus the above distance d (that can be either d_{HL} or d_{AD}) with a tuning parameter $a > 0$:

$$K_{\text{SEG}}(\mathbf{x}_i, \mathbf{x}_j) = 0.9^{a * d(\mathbf{x}_i, \mathbf{x}_j)} \tag{3}$$

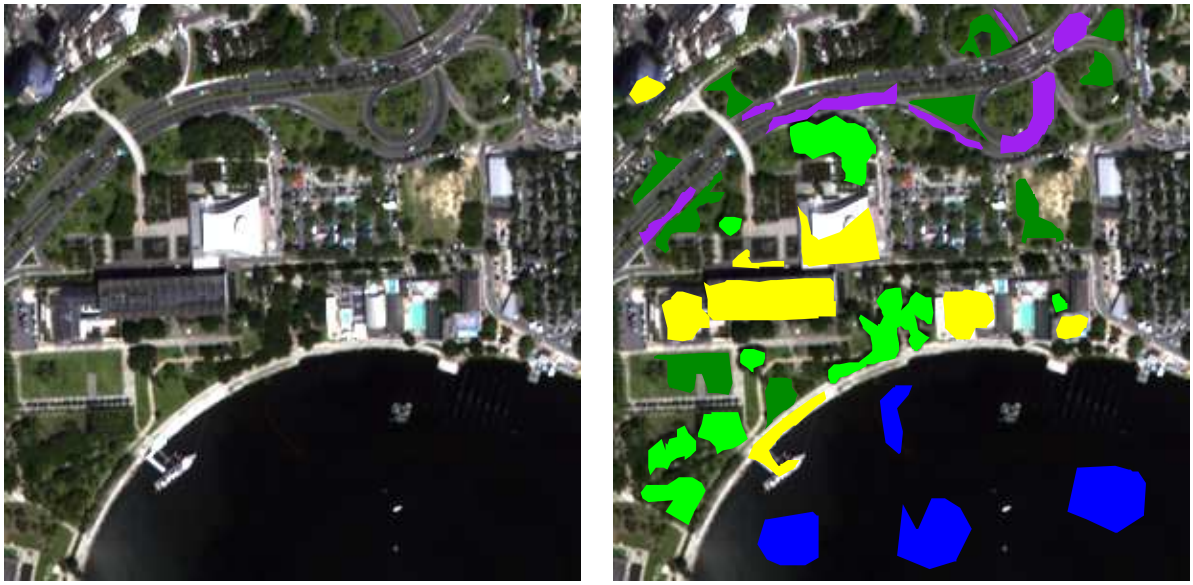
The parameter a controls the rate of degree of the similarity function: the higher this parameter, the steepest the exponential decrease of similarity when increasing the distance $d(\mathbf{x}_i, \mathbf{x}_j)$.

III. DATA SETS AND DESIGN OF EXPERIMENTS

A. Data Sets Used

Pavia dataset: airborne data from the ROSIS-3 (Reflective Optics System Imaging Spectrometer) hyperspectral optical sensor are used. The images were acquired over the city of Pavia, northern Italy. Due to noise effects, only 103 spectral channels are actually used. The image is 610 $\times$ 340 pixels, with a spatial resolution of 1.3m. Nine classes are of interest and defined in a ground truth composed of 42'776 labeled pixels (cf. Fig. 5 (a)).

Rio dataset: In this experiment, VHR multispectral data acquired by the WorldView-II sensor are used. The image covers a part of the center of Rio de Janeiro, Brazil [16]. Eight spectral



(a) (b)

Fig. 2. Rio dataset: (a) three bands composite image (bands 4, 3 and 2), (b) ground truth of the area (including water, tree, grass, road, and building).

bands are available at a spatial resolution of 2.0m. A 301 $\times$ 301 subset of the original nadir image is used for the experiments. Five classes are considered, for which a total of 15'294 labeled pixels is extracted by photointerpretation. The data are shown in Fig. 2.

B. Design of Experiments

Segmentation algorithm: to partition the image and obtain the hierarchy of segments, we used the HSEG algorithm [17]. The algorithm starts considering each pixel as a segment. Then the HSEG partitions the image iteratively, by merging the segments with minimal dissimilarity (typically the Euclidean or cosine distance between centroids of the segments) until all the image is assigned to a single segment. An important parameter in what follows is the *chk_nregions* parameter, which controls the number of regions in the most detailed level of the segmentation hierarchy returned by the algorithm (for which $h = 1$).

Classifier and free parameters: in our experiments, an one-against-all SVM was considered. As spectral kernel we used a RBF kernel, since it outputs similarities, which are positive and in the range $[0, 1]$, thus easing the combination with the other kernels without normalization of the values. The SVM parameters C and σ are tuned by a crossvalidation in the same range $\{10^{-3}$, ...

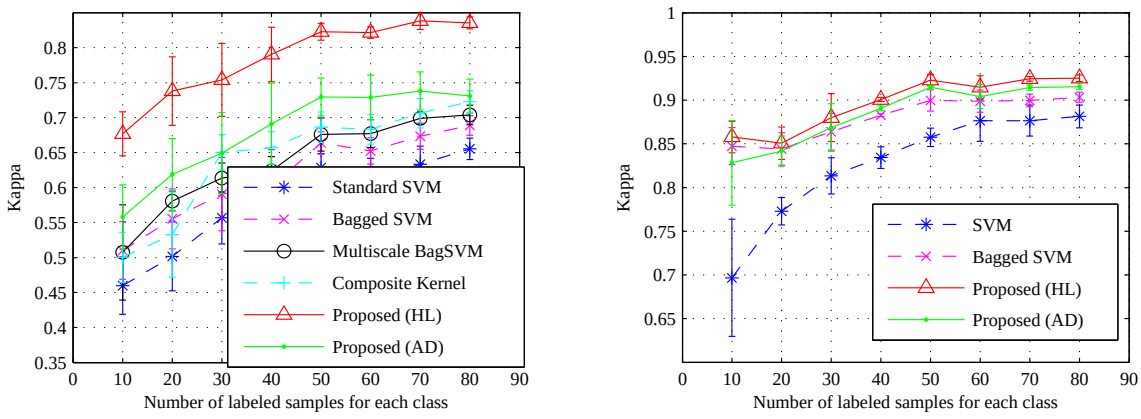
,10^3}. In the experiments, $u = 1000$ unlabeled pixels are randomly selected to construct K_{BAG}, as in [1]. The number of clusters k in the k-means algorithm used to build K_{BAG} is set to 30. Based on experimental analysis, the tuning parameter a in Eq. (3) is empirically set to 1 for the HL distance and to 50 for the AD method. Following a sensitivity study on different sizes of the labeled set, the weights for spectral kernel K_{RBF} and bagged kernel K_{BAG} are empirically set to $\alpha = 0.6$ and $\beta = 0.2$. The sensitivity study for these parameters is reported in section IV-B. The *chk_nregions* in the HSEG algorithm is set to 1024. The sensitivity study for *chk_nregions* is reported in section IV-C.

Assessment of performances: classification accuracy is evaluated in terms of both overall accuracy (OA, [%]) and Kappa statistic (κ). We test the proposed method for different amounts of randomly selected labeled samples with $n = \{10, \ldots, 80\}$ per class in the Pavia dataset and $n = \{5, \ldots, 40\}$ labeled pixels per class in the Rio dataset. For each set of the same size, 10 realizations with random training data sets are selected. The average Kappa obtained with these 10 random training sets are reported. We compare the results of the proposed kernel function with a standard RBF SVM, the Bagged SVM [1], the multiscale version of the bagged SVM in [8] and the spatio-spectral composite kernel [9] (which includes spatial regularization via spatial average filters). Finally, we also consider enriching the RBF kernel K_{RBF} with hierarchical spatial features, as in [14], thus showing the complementarity of the two approaches.

IV. Results and Analysis

A. Numerical Results

For the Pavia dataset (Fig. 3), the proposed method significantly improves Kappa of the Bagged SVM by 10%-15%. The spatial kernel using the HL distance performs better than the one issued from the aggregation distance. Besides, the proposed method performs better than the one obtained using the composite kernels. In all cases, the increases are significant, as tested by the McNemar's test. The better performance of HL distance can be explained as follows: using this distance, where each aggregation step denotes the same increase in distance, helps disentangling confusion for pixels that are from different classes but are spatially close and show similar spectral properties. The AD distance tends to return high K_{SEG} values for those pixels, since the distances in the lower part of the agglomeration hierarchy are very small. On the contrary, HL increases the dissimilarity between couples of pixels that merge in the lower



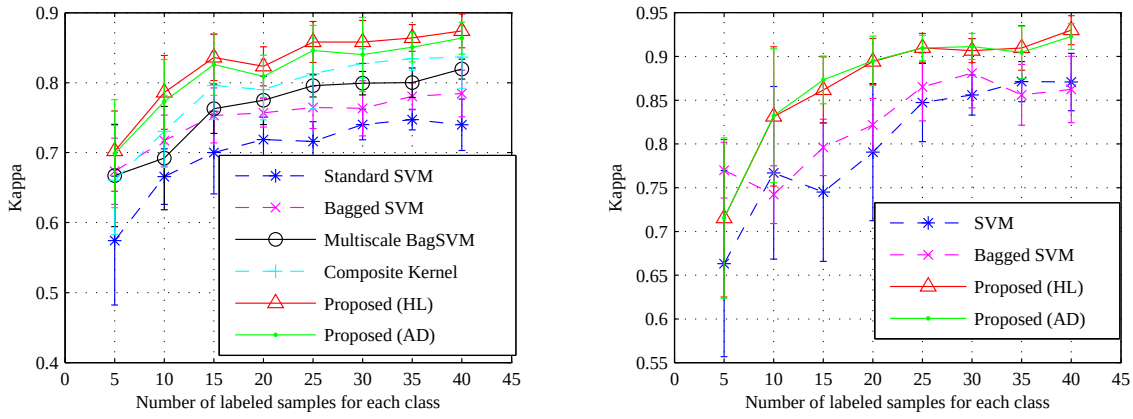
(a) (b)

Fig. 3. Classification accuracy of the proposed method for the Pavia dataset with K_{RBF} using spectral bands (a), and K_{RBF} exploiting multiscale contextual features (b), as in [14]. Bagged SVM is the method in [1], and Multiscale BagSVM is the method in [8]. Composite Kernel is the method in [9].

part of the hierarchy. When considering spatial features issued from hierarchical segmentation in the construction of K_{RBF} (as in [14]), the SVM baseline (blue dashed in Fig. 3(b)) improves by almost 20% and results in similar performances as the proposed method in the spectral case only (red line in Fig. 3(a)). But when such kernel is used in conjunction with the proposed K_{SEG} (red line in Fig. 3(a)), performances have an additional boost with respect to [14], comprised between 12% (10 labeled pixels per class) and 3% (80 labeled pixels per class).

As for the Rio dataset (Fig. 4), the proposed method improves Kappa of the Bagged SVM by 2%-7% depending on the labeled information available. We can explain the smaller increases of performance by the smaller complexity of this second dataset, which is of lower dimension ($d = 8$ against 103 in the Pavia case) and involves a smaller number of classes (5 against 9). Nonetheless, the two spatial kernels show a similar classification performance, but still outperform the competing methods.

Classification maps for the different methods for the Pavia dataset are shown in Fig. 5. Many parts of the meadow area (in green) in the image are misclassified as ”bare soil” class for the former three methods, especially for the lower part of the image, whereas most of these areas are correctly classified for the proposed method, especially for the one based on HL distance. Moreover, the proposed method provides more coherent classification maps than the



(a) (b)

Fig. 4. Classification accuracy of the proposed method for the Rio dataset with K_{RBF} using spectral bands (a), and K_{RBF} exploiting multiscale contextual features (b), as in [14]. Bagged SVM is the method in [1], and Multiscale BagSVM is the method in [8]. Composite Kernel is the method in [9].

other competing methods. This can be seen in particular for the ”Bare soil” class (in the center part of the image), where the maps obtained by the other methods are strongly contaminated by pixels misclassified as meadows (in green).

B. Sensitivity Analysis for the Kernel Weights

Fig. 6 shows the results obtained with different parameter weights α and β in the kernel combination: panel 6(a) shows the OA obtained when using a limited amount of training samples (10 samples per class), while panel 6(b) shows the results with more training data (80 samples per class). In both cases, the accuracies in the diagonal are lower than those in the off-diagonal elements (please note that the upper half of the triangle is empty, since no combinations with $\alpha + \beta > 1$ are considered). This diagonal corresponds to the case where $\alpha + \beta = 1$, which means that no spatial kernel has been used in the combination of Eq. (2). It also corresponds to the Bagged SVM of [1]. This shows even further the interest of adding the similarity measure depending on spatial information proposed in this paper.As observed in [10], the best combination is obtained when using all the kernels, as they encode different views on the classification problem: the spectral similarity (K_{RBF}), the structure of the manifold (K_{BAG}) and the spatial distribution (K_{SEG}). A source alone may result in suboptimal results: if using K_{SEG} only (case when $\alpha = \beta = 0$), the similarity becomes function of the spatial distribution

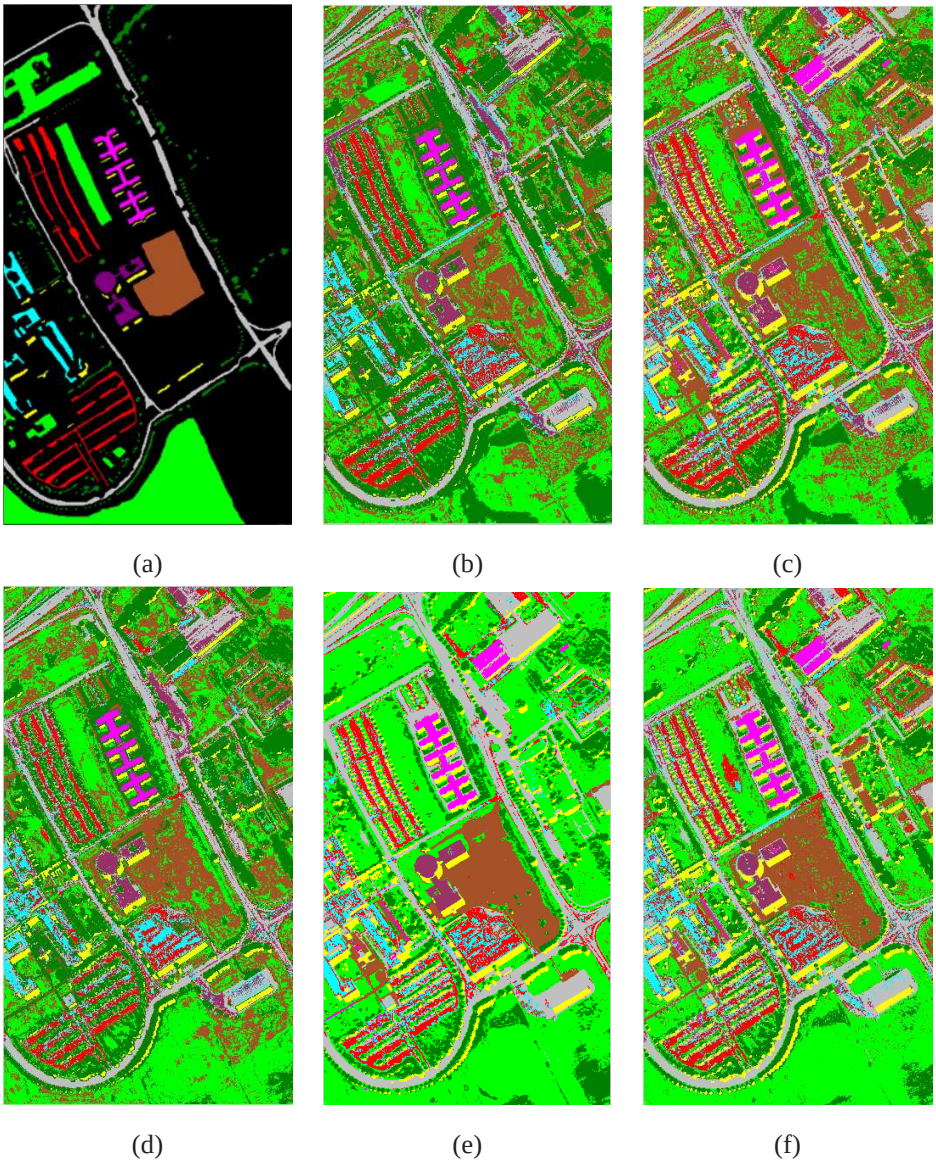
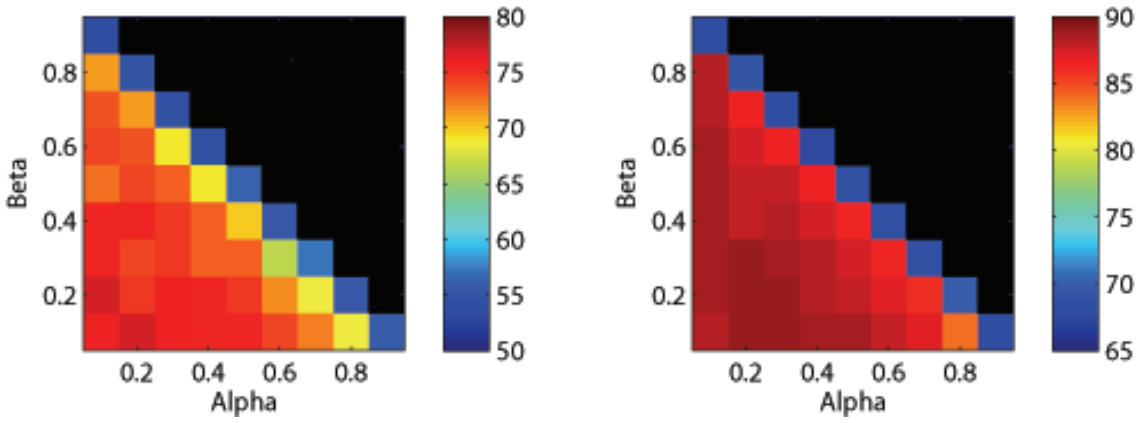


Fig. 5. Classification maps for the competing and proposed method for the Pavia dataset (a) ground truth, (b) Standard SVM, (c) Bagged SVM [1], (d) Composite kernel [9], (e) Proposed (Hierarchical level), and (f) Proposed (Aggregation distance).

of the training samples. If the samples are not distributed homogeneously, the quality of results will depend mainly on the proximity to such samples (in the Pavia experiments, using K_{SEG} only resulted in an OA between $62.6\% \pm 7.40$ (when $n = 40$) and $68.6\% \pm 3.04$ (when $n = 80$), while the spatio-spectral model proposed provides accuracies higher than 80% for both cases.

Similarly to what observed in [8], the weight for the K_{SEG} needs to be more important to obtain accurate classification results when only few labeled samples are available. With more



(a) (b)

Fig. 6. Overall Accuracy (OA,[%]) with respect to different kernel weights for Pavia dataset with (a) 10 training samples per class and (b) 80 training samples per class.

labeled samples, the effect of the spatial kernel is lower. This is due to the difficulty for the SVM to find the right level of regularization when using an insufficient amount of labeled pixels, difficulty that is partly resolved by the use of K_{BAG} (as shown in the diagonal of the plots in Fig. 6) and even further with the spatial regularization brought by K_{SEG}.

C. Results with Different Degrees of Spatial Regularization

The effects of the degree of spatial regularization are tested by varying the depth of the hierarchy obtained by the HSEG. By choosing different *chk_nregions* in the HSEG algorithm, we obtain a series of hierarchical results with different numbers of regions at the finest scale ($h = 1$), as shown in Fig. 7(b). The classification results using the spatial kernel derived from HL distance are shown in Fig. 7(a).

From Fig. 7(a) we can see that, with a very limited training set (10, 20 samples for each class), the results with a stronger spatial regularization (200, 400 segments) are better than those with weaker spatial regularization (600, 800 segments). This confirms the observations of the previous section, as it is reasonable to assume that, when considering an inaccurate classifier trained using a limited training set, a strong spatial regularization will strongly improve the classification results. When the number of training samples increase, the degree of spatial regularization becomes less crucial.

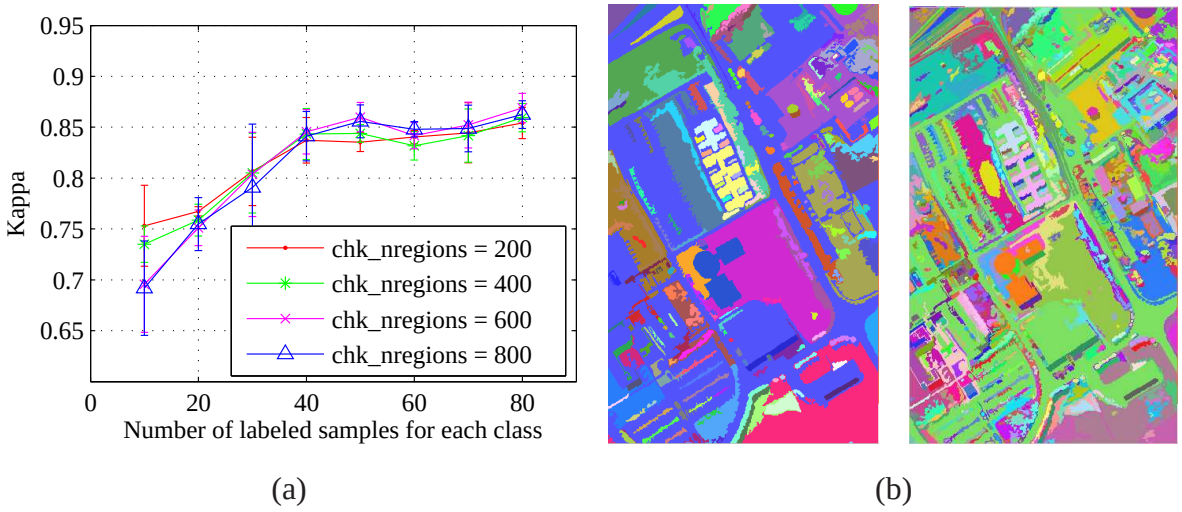


Fig. 7. (a) Kappa of the proposed method (Pavia dataset) with different settings of the HSEG algorithm (*chk_nregions* are the number of segments at the lowest level of aggregation, whose examples for 200 and 800 are reported in panel (b).

V. CONCLUSION

A kernel encoding spatial relationships between objects has been proposed for semisupervised remote sensing image classification. The spatial kernel is constructed from distances retrieved by a hierarchical segmentation. This kernel utilizes similarities observed at the pixel level and at the object level simultaneously, thus enforcing a spatially-dependent cluster assumption. Experimental tests demonstrate that the proposed method incorporating spatial similarity obtains higher classification accuracy and smoother classification maps than spectral classifiers and spatio-spectral methods based on composite kernels considering spatial filters.

ACKNOWLEDGEMENT

The authors would like to thank University of Pavia for providing the Pavia dataset, Digital Globe and the IEEE IADF TC for making the Rio dataset available. The authors also would like to thank three reviewers for their suggestions.

REFERENCES

[1] D. Tuia and G. Gamps-Valls, "Semisupervised remote sensing image classification with cluster kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 224–228, Apr. 2009.

[2] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning, 1st ed.* Cambridge: MA:MIT Press, 2006.

[3] Q. Jackson and D. A. Landgrebe, "An adaptive classifier design for high-dimensional data analysis with a limited training data set," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, pp. 2664–2679, 2001.

[4] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive svm for the semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[5] G. Gamps-Valls, T. V. B. Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.

[6] L. Gómez-Chova, G. Gamps-Valls, J. Muñoz-Marí, and J. Calpe-Maravilla, "Semisupervised image classification with Laplacian support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 3, pp. 336–340, Jul. 2008.

[7] J. Weston, C. Leslie, E. Ie, D. Zhou, A. Elisseeff, and W. S. Noble, "Semisupervised protein classification using cluster kernels," *Bioinformatics*, vol. 21, no. 15, pp. 3241–3247, Aug. 2005.

[8] D. Tuia and G. Gamps-Valls, "Urban image classification with semisupervised multiscale cluster kernel," *IEEE J. Sel. Topics Applied Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 65–74, Mar. 2011.

[9] G. Camps-Valls, L. Gómez-Chova, J. Muoz-Marí, 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, 2006.

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

[11] M. Marconcini, G. Gamps-Valls, and L. Bruzzone, "A composite semisupervised SVM for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 234–238, Apr. 2009.

[12] L. Capobianco, A. Garzelli, and G. Camps-Valls, "Target detection with semisupervised kernel orthogonal subspace projection," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3822–3833, Jul 2009.

[13] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[14] L. Bruzzone and L. Carlin, "A multilevel context-based system for classification of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 9, pp. 2587–2600, 2006.

[15] J. Muñoz-Marí, D. Tuia, and G. Camp-Valls, "Semisupervised classification of remote sensing images with active queries," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3751–3763, Oct. 2012.

[16] F. Pacifici and Q. Du, "Foreword to the special issue on optical multiangular data exploitation and outcome of the 2011 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 3–7, 2012.

[17] J. Tilton, "RHSEG user's manual: Including the core RHSEG Open Source Release, HSEGExtract, HSEGReader and HSEGViewer (Version 1.47)," http://techtransfer.gsfc.nasa.gov/ft_tech_rhseg.shtm, 2009.