Multisource Composite Kernels for Urban-Image Classification

Devis Tuia, *Student Member, IEEE*, Frédéric Ratle, Alexei Pozdnoukhov, and Gustavo Camps-Valls, *Senior Member, IEEE*

***Abstract*—This letter presents advanced classification methods for very high resolution images. Efficient multisource information, both spectral and spatial, is exploited through the use of composite kernels in support vector machines. Weighted summations of kernels accounting for separate sources of spectral and spatial information are analyzed and compared to classical approaches such as pure spectral classification or stacked approaches using all the features in a single vector. Model selection problems are addressed, as well as the importance of the different kernels in the weighted summation.**

***Index Terms*—Multiple kernel learning, support vector machines (SVMs), urban monitoring, very high resolution image.**

I. Introduction

VERY high spatial resolution (VHR) has been one of the major achievements of satellite imagery in the last decades. Sensors providing submetric resolution have been developed, and satellites such as QuickBird, GeoEye-1, or Pléiades have been or are about to be launched. These sensors provide images that are unique in terms of spatial detail and open a wide range of challenges for geospatial information processing.

The classification of land use at such a high resolution is a very challenging problem, because, even if it provides more details about the boundaries and shape of the objects, VHR also introduces undesired small-scale objects, such as chimneys or cars. Therefore, the integration, or fusion, of several sources of information capable to discriminate the classes of interest and to filter nonrelevant information becomes crucial. For instance, the simultaneous availability of panchromatic and multispectral optical imagery at submetric resolution permits one to design classifiers exploiting both spectral and spatial information. By spatial information, we mean the integration, at the pixel level, of the spatial configuration of the objects of the scene. Such an integration can be achieved by using textural or morphological features. In this contribution, we focus on morphological features.

Mathematical morphology [1] provides a collection of image filters (called operators) that analyze the image with respect to the distribution of gray levels in the spatial neighborhood of the pixels. In optical imagery, morphological operators have been used to classify images at high resolution and have been highlighted as very promising tools for data analysis [2], [3]. Tuia *et al.* applied morphological indexes to QuickBird imagery for urban classification: in [4] using only panchromatic images and in [5] using spectral and spatial information with a multiscale support vector machine (SVM) classifier. Both works showed the interest of using morphological filters for VHR urban classification. In [5], the properties of composite kernels (see [8]) have been used to build a multiscale classifier exploiting simultaneously spectral and spatial information. In this study, each scale was represented by a kernel applied to the complete feature set, i.e., the stacked (concatenated) vector containing all the spectral/spatial features. Therefore, despite its good performance, the contribution of spatial features and the concept of scales are difficult to interpret and quantify.

In this letter, the combination of kernels accounting separately for spectral and spatial information are studied and discussed. This approach is closer to the one proposed in [6] for panchromatic images, in [7]–[9] for hyperspectral imagery, in [10] for synthetic aperture radar (SAR) images, and in [11] for multitemporal and multisource optical and SAR images. In these works, each source of information is handled by a single kernel, and the optimal weighted combination is found. The novelty of the present work is found in 1) the application of composite kernels to urban VHR imagery and 2) the detailed analysis of the relative relevance of different information sources. The first goal is motivated by the new problems posed with the higher spatial resolution of the images. The analysis of the composite kernels is tackled by optimizing the weights in the linear combination and by exploring generalization bounds of performance.

The remainder of this letter is organized as follows. Section II illustrates the principles of both SVMs and composite kernels. Section III shows the data sets used as well as the experiments, whose results are discussed in Section IV.

II. Composite Kernels for SVM

This section presents the composite kernel theory and the properties of Mercer's kernels necessary to build valid

Manuscript received September 30, 2008; revised November 12, 2008. First published May 5, 2009; current version published January 13, 2010. This work was supported by the Swiss National Foundation under Grant 100012-113506, Grant 105211-107862, and Grant 200020-121835, by the Spanish Ministry of Science under Grant CONSOLIDER/CSD 2007-0018 and Grant AYA2008-05965-C04-3, and by the Science Foundation Ireland under the National Development Plan through Strategic Research Cluster Grant 07/SRC/I1168.

D. Tuia and F. Ratle are with the Institute of Geomatics and Analysis of Risk, University of Lausanne, 1015 Lausanne, Switzerland (e-mail: devis.tuia@unil.ch; frederic.ratle@unil.ch).

A. Pozdnoukhov was with the Institute of Geomatics and Analysis of Risk, University of Lausanne, 1015 Lausanne, Switzerland. He is now with the National Centre for Geocomputation, National University of Ireland Maynooth, Ireland (e-mail: alexei.pozdnoukhov@nuim.ie).

G. Camps-Valls is with the Image Processing Laboratory, Escola Tècnica Superior d'Enginyeria, Universitat de València, 46100 València, Spain (e-mail: gustavo.camps@uv.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2009.2015341

composite kernels. A brief reminder about SVMs and kernel methods is also given.

A. SVMs and Kernels

The SVM is one of the most famous kernel methods. The SVM is a classifier that finds the hyperplane which is the farthest away from the closest training samples. The distance between the hyperplane and the closest training points is called the margin, which is the quantity maximized by the SVM.

Consider a linearly separable case and a set of n labeled samples $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and associated labels $Y = \{y_1, y_2, \ldots, y_n\} \in \{-1, +1\}$. We are searching for a linear decision function $f(x) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ maximizing the margin. It can be shown [12] that minimizing the norm of the parameters $1/2\|\boldsymbol{w}\|^2$ under the constraint $y_i(\langle \mathbf{w}, \mathbf{x} \rangle + b) \geq 1$ maximizes the margin. Such a minimization of the weights provides a naturally regularized solution, which favors smooth models of optimal complexity and avoids overfitting the data. The SVM dual problem can be written as

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \right\}. \tag{1}$$

The solution of this problem is found after the maximization of (1)

$$\langle \mathbf{w}, \mathbf{x} \rangle = \sum_{j=1}^{n} y_j \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \tag{2}$$

The support vectors, i.e., the points lying on the margin, are the samples with associated coefficients $\alpha_i \neq 0$.

Kernel methods [13] implicitly map the input space $\mathcal{X}$ of the original (and often nonlinearly separable) data into a higher dimensional Hilbert space $\mathcal{H}$, through a kernel function K that represents a dot product in $\mathcal{H}$. It is assumed that the data are more likely to be linearly separable in $\mathcal{H}$. We should note that both the SVM problem (1) and the solution (2) solely depend on the similarity between training patterns, not on the samples themselves. Thus, since a kernel function K represents a dot product in $\mathcal{H}$, it can replace the similarity $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$. This way, a linear hyperplane is found in the higher dimensional Hilbert space defined by the kernel, and now, the problem becomes

$$\max_{\alpha_i} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\}. \tag{3}$$

In order to predict an unknown sample $\mathbf{z}_i$, the following expression is used:

$$f(\mathbf{z}_i) = \text{sign} \left(\sum_{j=1}^{n} y_j \alpha_j K(\mathbf{x}_j, \mathbf{z}_i) + b \right). \tag{4}$$

B. Composite Kernels

Kernel matrices encode the similarity between points using a metric defined by the type of kernel function used. For instance, a linear kernel computes the similarity using a simple dot product $K(\mathbf{x}, \mathbf{z}) = \langle \mathbf{x}, \mathbf{z} \rangle$. Other usual kernels are the polynomial kernel $K(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle + 1)^d$ and the radial basis function kernel $K(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2)/2\sigma^2$.

Some properties of Mercer's kernels that are relevant for this letter are as follows.

Property 1: Let K_1 and K_2 be two Mercer's kernels on $\mathcal{X} \times \mathcal{X}$ and $\mu > 0$. Then, the kernels

$$K(\mathbf{x}, \mathbf{z}) = K_1(\mathbf{x}, \mathbf{z}) + K_2(\mathbf{x}, \mathbf{z}) \tag{5}$$
$$K(\mathbf{x}, \mathbf{z}) = K_1(\mathbf{x}, \mathbf{z}) \cdot K_2(\mathbf{x}, \mathbf{z}) \tag{6}$$
$$K(\mathbf{x}, \mathbf{z}) = \mu K_1(\mathbf{x}, \mathbf{z}) \tag{7}$$

are valid Mercer's kernels.

Therefore, one can design kernels by direct sum or product of (weighted) valid kernels.

Note that the use of composite kernels does not increase the computational cost of the SVM. Once all kernels are computed and combined, the composite kernel is fed into a standard SVM solver [15].

III. Methods and Data

In this letter, we want to evaluate the impact of different sources of information using a multisource composite kernel. Three groups of features have been considered:

1) MS: the four multispectral bands of a QuickBird image;
2) OR: nine morphological features extracted from a panchromatic QuickBird image using an opening-by-reconstruction morphological filter [1];
3) CR: nine morphological features extracted by applying the closing by reconstruction filter.

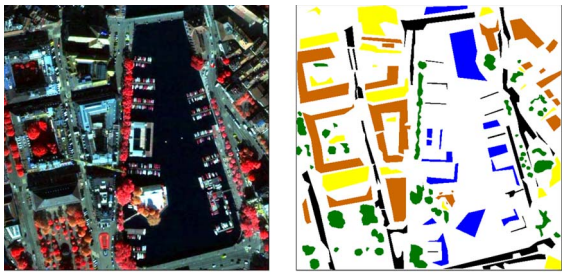
OR filters the pixels that are brighter than their surroundings. CR produces the same effect for darker pixels. All the morphological images have been extracted using a diamond-shaped structuring element with an increasing radius going from 3 to 19 pixels. Each feature has been individually reduced to standard scores.

A. Experiments

Six different experiments have been carried out, accounting for different combinations of models. A radial basis function (RBF) kernel has been used in all the experiments. In the following, each kernel encodes the similarity between a separate set of features, respectively, MS, OR, and CR. We will refer to these kernels as $K^{\text{MS}} = K(\mathbf{x}^{\text{MS}}, \mathbf{x}^{\text{MS}})$, $K^{\text{OR}} = K(\mathbf{x}^{\text{OR}}, \mathbf{x}^{\text{OR}})$, and $K^{\text{CR}} = K(\mathbf{x}^{\text{CR}}, \mathbf{x}^{\text{CR}})$.

The experiments are grouped depending on the type of kernel composition used.

1) Experiments 1–4 use a single-kernel SVM.
   - Experiments 1–3. *Single groups* (MS, OR, and CR): A single group of bands is used separately with its specific (dedicated) kernel (K^{MS}, K^{OR}, or K^{CR}). The kernel parameter has been tuned in the range $\sigma = \{0.01, \ldots, 0.5\}$, and the cost parameter has been tuned in the range $C = \{1, \ldots, 75\}$.
   - Experiment 4. *Stacked approach*: All the features are stacked in a 22-dimensional vector $\mathbf{x}^{\text{ST}} = \{\mathbf{x}^{\text{MS}}, \mathbf{x}^{\text{OR}}, \mathbf{x}^{\text{CR}}\}$. A single-scale kernel $K^{\text{ST}} = K(\mathbf{x}^{\text{ST}}, \mathbf{x}^{\text{ST}})$ is used.
2) Experiments 5 and 6 use composite summation kernels built by exploiting the properties shown in Property 1.



(a) (b)

Fig. 1. (a) Multispectral image and (b) ground survey used [Green = trees (T); blue = water (W); orange = buildings (B); black = roads (R); yellow = shadows (S)].

Note that we restrict ourselves to summation kernels, and no cross-kernels such as $K(\mathbf{x}^{\mathrm{MS}}, \mathbf{x}^{\mathrm{OR}})$ are used. In principle, the number of features for each source could be different, making its use impossible. In addition, previous work [7], [11] demonstrated that simpler kernels yielded better results in general.

The weighted summation of kernels is defined as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \alpha K\left(\mathbf{x}_i^{\mathrm{MS}}, \mathbf{x}_j^{\mathrm{MS}}\right) + \beta K\left(\mathbf{x}_i^{\mathrm{OR}}, \mathbf{x}_j^{\mathrm{OR}}\right) + \gamma K\left(\mathbf{x}_i^{\mathrm{CR}}, \mathbf{x}_j^{\mathrm{CR}}\right) \quad (8)$$

where $\{\alpha, \beta, \gamma\} \geq 0$ and $\alpha + \beta + \gamma = 1$, as in [14]. This way, the resulting kernel is normalized, and the relative importance of each type of information is directly obtained from the optimization.

Experiment 5. *Sum Models*: The three kernels' free parameters are optimized separately (in the MS, OR, and CR experiments) before the optimization of the combination of kernels, which is computed by optimizing α, β, and γ only. This way, we try to limit the size of the parameter space to be searched.

Experiment 6. *Multioptimization*: The free parameters associated to each RBF kernel (σ^{MS}, σ^{OR}, and σ^{CR}) are optimized simultaneously with α, β, and γ with a cross-validation procedure.

A multiclass SVM (with the one-against-all approach) handling composite kernels has been implemented using the Torch 3 library [15]. A C# application has been developed to extract morphological images.

The free parameters are tuned during training by grid search over the parameter space $\Phi = \{\sigma^{\mathrm{MS}}, \sigma^{\mathrm{OR}}, \sigma^{\mathrm{CR}}, C, \alpha, \beta, \gamma\}$, where σ^{MS}, σ^{OR}, and σ^{CR} are the bandwidths of the single kernels. C is a regularization parameter controlling the generalization capabilities of the classifier.

B. Data Set

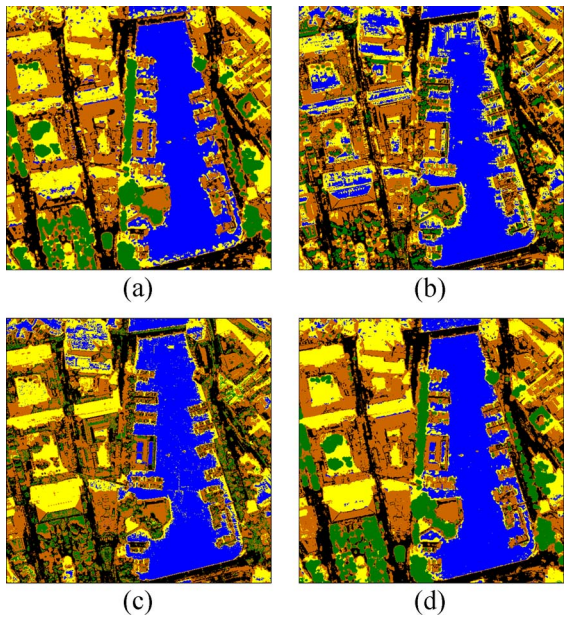
Experiments have been conducted on a QuickBird image of the city of Zürich, Switzerland. The image [Fig. 1(a)] was taken on October 17, 2006. Five classes of interest were extracted by the visual inspection of the image, and a ground truth of 105 011 pixels [Fig. 1(b)] was created. Table I illustrates the classes, the colors used in the graphics, and the random data split in three subsets: training (4000 pixels), testing (400 pixels used for model selection), and validation (the remaining 97 011 pixels).

TABLE I
GROUND TRUTH USED IN THE EXPERIMENTS

Class	Color	Training	Testing	Validation
Trees (T)	green	660	664	15318
Water (W)	blue	564	567	14014
Building (B)	orange	1253	1286	30664
Roads (R)	black	828	835	20143
Shadow (S)	yellow	696	648	16873
Total		4000	4000	97011

TABLE II
CLASSIFICATION ACCURACIES (IN PERCENT) AND KAPPA STATISTICS FOR THE SIX EXPERIMENTS CONSIDERED

	MS	OR	CR	Stacked	Sum Models	Multiopt
Trees	**86.08**	32.01	20.38	93.20	89.12	**94.18**
Water	90.00	87.88	**95.74**	**96.90**	94.88	96.87
Building	**83.79**	69.74	63.35	83.43	**83.87**	83.83
Road	**81.86**	80.91	72.88	87.08	83.28	**89.38**
Shadow	**91.85**	53.36	74.91	**96.34**	94.48	95.64
Overall	**86.05**	65.87	65.23	89.92	88.01	**90.55**
Kappa Index	**0.82**	0.56	0.55	0.87	0.85	**0.88**



(a) (b) (c) (d)

Fig. 2. Classification maps using (a) MS, (b) opening by reconstruction, (c) closing by reconstruction, and (d) multioptimization composite kernels.

IV. RESULTS AND DISCUSSION

This section is devoted to the analysis of the results obtained within the aforementioned experimental framework. Furthermore, free parameter selection is addressed through novel multioptimization maps, and some recommendations for further use are devised.

Table II shows the classification results for the six experiments, while the classification maps are shown in Fig. 2. The benefits of adding spatial information to the generalization capabilities of the SVM is evaluated using the $\xi\alpha$-estimator [16], which evaluates the upper bound of the SVM error as follows:

$$Err^n_{\xi\alpha}(\mathrm{SVM}) = \frac{d}{n}, \qquad \text{with } d = |\{i : (2\alpha_i + \xi_i) \geq 1\}|. \quad (9)$$

Regarding the single groups, the "MS" experiment provides the best results: a classifier exploiting the spectral information only results in an overall accuracy of 86.05% with a Kappa

TABLE III
CONFUSION MATRICES FOR THE VALIDATION OF THE MS, OR, CR, AND MULTIOPTIMIZATION EXPERIMENTS. IN BOLD ARE THE MAIN ELEMENTS DISCUSSED IN THE TEXT. (%UA = USER'S ACCURACY; %PA = PRODUCER'S ACCURACY)

	T	W	B	R	S	%UA
T	13186	1	**1766**	21	344	97.17
W	0	12613	160	0	**1242**	94.01
B	190	206	25695	**3130**	1444	80.73
R	15	3	**3610**	16490	26	83.93
S	179	594	597	6	15498	83.53
%PA	86.08	89.99	83.79	81.86	91.85	

MS

	T	W	B	R	S	%UA
T	4903	7	5440	4415	553	81.83
W	0	12317	62	0	1636	61.05
B	513	281	21385	6664	1822	70.38
R	573	3	3241	16298	29	59.44
S	3	7568	255	44	9004	69.03
%PA	32.01	87.88	69.74	80.91	53.36	

OR

	T	W	B	R	S	%UA
T	3122	0	6975	4747	474	41.84
W	0	13418	19	0	**578**	77.93
B	2967	89	19427	6981	1201	62.73
R	1369	0	4092	14680	3	55.46
S	4	3712	455	62	12641	84.86
%PA	20.38	95.74	63.35	72.87	74.91	

CR

	T	W	B	R	S	%UA
T	14427	0	**542**	40	309	97.94
W	0	13576	23	2	**414**	96.85
B	165	59	25705	3666	1070	89.86
R	93	0	**2043**	18005	3	82.87
S	45	382	294	14	16139	89.99
%PA	94.18	96.87	83.83	89.38	95.64	

Multiopt.

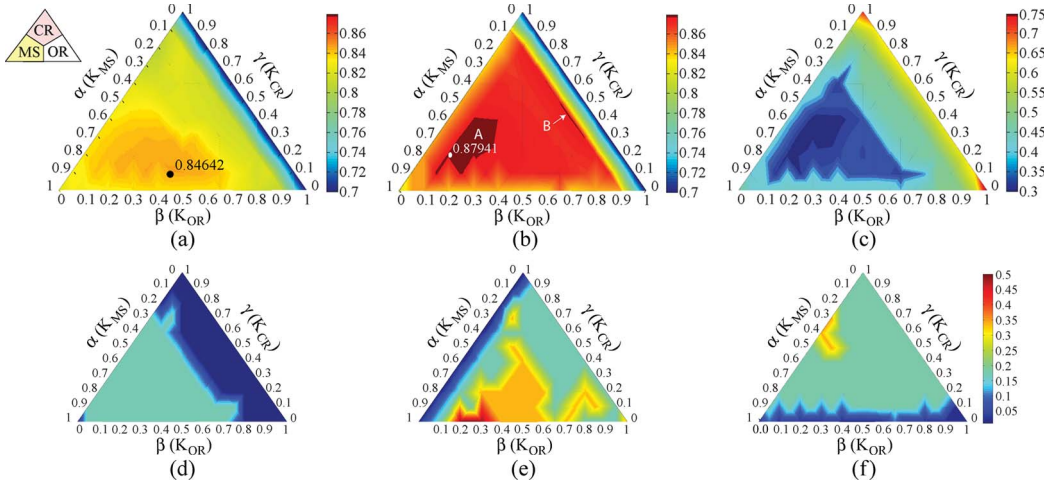


Fig. 3. Surface of the influence of the α, β, and γ parameters on the Kappa index for the (a) Sum Models ($\sigma^{\mathrm{MS}} = \sigma^{\mathrm{OR}} = \sigma^{\mathrm{CR}} = 0.01$) and (b) multioptimization approach. The areas of influence of the features are shown in the small triangle on the top left. (c) $\xi\alpha$ error bound for the multioptimization approach (value in the corners are the bounds for the single experiments MS, OR, and CR). The optimal values of the multioptimization experiment are shown in (d) for σ^{MS}, (e) for σ^{OR}, and (f) for σ^{CR}: Blue areas correspond to areas where the optimal parameter is the same as in the experiments using a single feature set. Each point in (b) has been obtained by a model using the optimal sigmas reported at the same point in triangles (d), (e), and (f).

index of 0.82. The classification map [Fig. 2(a)] shows a general correct reconstruction of the objects of the scene, although confusion between buildings and, respectively, roads and trees is visible (see also the confusion matrix given in Table III—MS). Confusion between shadows and waterplans can be seen at the borders of the roofs for the class Shadow and at the southern and northern ends of the waterplan for the class Water.

Poor results are provided by the "OR" and "CR" experiments that use morphological features independently (Kappa indexes of 0.55 and 0.56). Nonetheless, some observations rise from these results: First, OR features equal the result of MS features for the class Roads (see Fig. 2(b), for instance, the bridge on top of the image). Second, CR provides the best result for the class Water (see the confusion matrix of Table III—CR) and can solve the ambiguity between water and shadow in the lower half of the classification map in Fig. 2(c). Summing up, the morphological images construct good kernels for classes having very similar spectral response, but being characterized by different shapes of the objects.

The simultaneous use of all the features in the "Stacked" experiment leads to an improvement of the MS result of about 4% (0.05 for Kappa index). A SVM using a single kernel tuned on all the features (optimal σ is found at 0.173) provides a good result, where only a small confusion between vegetation and buildings can be noticed. Confusion between roads and buildings is no longer present (apart from atypical pixels such as details of the roofs), and errors between water and shadows also disappear. Therefore, the inclusion of the morphological features results in an improved classification of the scene. The "Sum Models" experiment was meant to explore the possibility to avoid the optimization of the kernel free parameters σ^{MS}, σ^{OR}, and σ^{CR}. The combination is done by using the optimal parameters found during the previous experiments using single kernels. This results in the parameter set $\sigma^{\mathrm{MS}} = \sigma^{\mathrm{OR}} = \sigma^{\mathrm{CR}} = 0.01$. Unfortunately, this experiment shows the worst result for the sets using multisource information and results in an overall accuracy of 88% with a related Kappa of 0.85. The classes Trees and Roads show the most important decrease in accuracy, certainly related to kernels using too narrow bandwidths. Fig. 3(a) shows the surface of Kappa statistics for all the combinations of parameters $\{\alpha, \beta, \gamma\}$: The maximum is found for $\{\alpha = 0.5, \beta = 0.4, \gamma = 0.1\}$, i.e., a kernel giving the most importance to the multispectral bands and using the OR information. The area showing the maximum values of Kappa is located in the area of influence of K^{MS}, and best results are obtained when giving a bigger importance to K^{OR} than to K^{CR}.

With an overall accuracy of 90.55% and a related Kappa index of 0.88, the "Multioptimization" experiment provides the best results. The update of the single-kernel parameters allows the algorithm to converge to a more suitable solution. By looking at the σ surfaces of Fig. 3(c), the optimal parameters σ^{MS}, σ^{OR}, and σ^{CR} are not the same as in the experiments using the single kernels. All the single kernels are related to an optimal σ of 0.01 [blue areas in Fig. 3(c)], while the reoptimization of the parameters lead to higher values of these parameters. Looking at the resulting surface of the Kappa statistic [Fig. 3(b)], we can observe that the models take advantage of the reestimation for every combination of kernel weightings $\{\alpha, \beta, \gamma\}$. Moreover, the area showing the best results (marked by A in the figure) is again the area of influence of K^{MS}, but it shows a better equilibrium between K^{OR} and K^{CR}: Optimal values are found for $\{\alpha = 0.7, \beta = 0.1, \gamma = 0.2\}$. A second area showing quasi-optimality is the one marked by B in Fig. 3(b): This area is linked to weighting parameters $\{\alpha = 0.2, \beta = 0.4, \gamma = 0.4\}$ and shows a situation where the spectral information is used only a little and the classification is achieved using the morphological information only.

The classification map given in Fig. 2(d) shows the best reconstruction of the scene. Note, in particular, the tree line at the right of the image and the good classification of the roads. The better separation between buildings and trees can also be seen in the confusion matrix of Table III—Multiopt., where the errors between these two classes decrease from 1766 pixels (MS) to 542. The same holds for the classes Water and Shadow, for which the error is reduced from 1242 to 414 pixels.

Regarding generalization capabilities (9), the experiments using single groups of features achieve $\xi\alpha$-error bounds of 42.5% when using MS bands and 68.18% and 75.05% when using the OR and CR bands, respectively. Therefore, an SVM trained using multispectral bands is more likely to achieve better generalization than one using spatial information only. Nevertheless, using both sets of information simultaneously results in a decrease of the $\xi\alpha$-bound, equal to 30.53% for the "Stacked" experiment and to 29.7% for the Multioptimization [see Fig. 3(c)]. These results confirm that the multioptimization approach achieves the best generalization properties for all the considered models.

V. Conclusion

In this letter, SVMs using composite kernels have been studied for the classification of very high resolution urban images. Morphological features extracted from the panchromatic image and multispectral bands have been used simultaneously to account for both spectral and spatial information. In order to assess the effect of each kind of features, kernels using independent information have been trained and combined using weighted summations.

The experiments made have shown a significant increase of the classification accuracy when the spatial information is used, confirming the results of [5], [7], and [11] and providing an easy-to-interpret look at the role of the different features for the final result.

Moreover, the results obtained using multisource information have shown an improvement of the generalization capabilities of the SVM, assessed through the estimation of the $\xi\alpha$ upper error bound: Kernels using spatial and spectral information led to a better generalization of the SVM, and composite kernels yielded the best results.

Nonetheless, such a detailed study is not operational for real classification tasks, where the whole parameter space cannot be searched in order to find the optimal parameters. Since traditional optimization methods such as steepest gradient descent cannot handle this kind of problems involving a complex parameter space, efficient optimization methods have to be considered and tested. Methods such as the multiple kernel learning chunking algorithm [14] will be the interest of future research.

Acknowledgment

The authors would like to thank D. Fasbender (Catholic University of Louvain) for his help with the graphics.

References

[1] P. Soille, *Morphological Image Analysis*. Berlin, Germany: Springer-Verlag, 2004.

[2] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.

[3] M. Fauvel, J. Chanussot, J. A. Benediktsson, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2007, pp. 4834–4837.

[4] D. Tuia, F. Pacifici, A. Pozdnoukhov, C. Kaiser, D. Solimini, and W. J. Emery, "Very-high resolution image classification using morphological operators and SVM," in *Proc. IGARSS*, 2008, pp. IV-220–IV-223.

[5] D. Tuia and F. Ratle, "Mixed spectral-structural classification of very high resolution images with summation kernels," in *Proc. SPIE Remote Sens. Conf.*, Cardiff, U.K., 2008, pp. 710 906-1–710 906-10.

[6] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A joint spatial and spectral SVMs classification of panchromatic images," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2007, pp. 1497–1500.

[7] G. Camps-Valls, L. Gómez-Chova, J. Munoz-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, Jan. 2006.

[8] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Adaptive pixel neighborhood definition for the classification of hyperspectral images with support vector machines and composite kernel," in *Proc. IEEE ICIP*, 2008, pp. 1884–1887.

[9] A. Villa, M. Fauvel, J. Chanussot, P. Gamba, and J. A. Benediktsson, "Gradient optimization for multiple kernel's parameters in support vector machines classification," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2008, pp. IV-224–IV-227.

[10] G. Mercier and F. Girard-Ardhuin, "Partially supervised oil-slick detection by SAR imagery using kernel expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2839–2846, Oct. 2006.

[11] G. Camps-Valls, L. Gómez-Chova, J. Munoz-Marí, J. L. Rojo-Álvarez, and M. Martínez-Ramón, "Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 6, pp. 1822–1835, Jun. 2008.

[12] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th ACM Workshop Comput. Learn. Theory*, 1992, pp. 144–152.

[13] F. Schweitzer and J. Steinbink, *Fundamental Principles of Urban Growth, Chapter Analysis and Computer Simulation of Urban Cluster Distributions*. Wuppertal, Germany: Müller + Busmann, 2002, pp. 142–157.

[14] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, vol. 7, pp. 1531–1565, Dec. 2006.

[15] R. Collobert, S. Bengio, and J. Mariéthoz, "Torch: A modular machine learning software library," IDIAP Res. Inst., Martigny, Switzerland, Tech. Rep. RR 02-46, 2002.

[16] T. Joachims, "Estimating the generalization performance of an SVM efficiently," in *Proc. 17th Int. Conf. Mach. Learn.*, 2000, pp. 431–438.