Design of a Spectral–Spatial Pattern Recognition Framework for Risk Assessments Using Landsat Data—A Case Study in Chile

Andreas Christian Braun, Carolina Rojas, Cristian Echeverri, Franz Rottensteiner, Hans-Peter Bähr, Joachim Niemeyer, Mauricio Aguayo Arias, Sergey Kosov, Stefan Hinz, and Uwe Weidner

***Abstract*—For many ecological applications of remote sensing, traditional multispectral data with moderate spatial and spectral resolution have to be used. Typical examples are land-use change or deforestation assessments. The study sites are frequently too large and the timespan covered too long assumes the availability of modern datasets such as very high resolution or hyperspectral data. However, in traditional datasets such as Landsat data, separability of the relevant classes is limited. A promising approach is to describe the landscape context pixels that are integrated. For this purpose, multiscale context features are computed. Then, spectral–spatial classification is employed. However, such approaches require sophisticated processing techniques. This study exemplifies these issues by designing an entire framework for exploiting context features. The framework uses kernel-based classifiers which are unified by a multiple classifier system and further improved by conditional random fields. Accuracy on three scenarios is raised between 19.0%pts and 26.6%pts. Although the framework is designed, focusing an application in Chile, it is generally enough to be applied to similar scenarios.**

***Index Terms*—Conditional random fields (CRFs), extended morphological profiles (EMPs), import vector machines (IVM), kernel composition, support vector machines (SVMs).**

I. Introduction

THE coastal range of central Chile has experienced as one of the fastest and most widespread land-use change processes worldwide [21]. During the last 35 years, establishment of commercial tree plantation forestry—heavily fostered by government subsidies—caused a worrisome process of deforestation of native forests. A range of ecological risks can arise from this process, among the biggest being the loss of biodiversity [46].

In order to analyse this process and its consequences, precise and detailed land-use maps are required. For ecological studies on deforestation, the accuracy of such maps should be higher than 90% [45]. This requirement is especially challenging since plantations and native forests both represent closed canopy tree ecosystems. Thus, they are spectrally similar and hard to separate.

In order to produce maps that distinguish spectrally similar vegetation classes, very high-resolution or hyperspectral remote-sensing data are frequently used [25], [37].

However, given the spatial extent of the coastal range and the fact that the timespan since 1974 has to be covered, such data are not available. Due to these circumstances, less modern multispectral data from Landsat have to be used—a common drawback to long-term ecological monitoring tasks, especially in developing countries [17].

However, due to the limited number of channels and the moderate resolution, distinguishing spectrally similar classes in Landsat data is challenging. Based on using spectral information only, the required accuracy level can hardly be achieved when using Landsat data [50].

A possible solution is to exploit the landscape context. Objects of the study site (e.g., trees) are integrated into a landscape context (e.g., open soils between the trees). This landscape context creates textures in the image. A plantation may exhibit a texture created through the regular planting pattern which helps to overcome the limits of spectral similarity when distinguishing it from a native forest. Therefore, computing context features from the spectral features in the image that describe the landscape context at the study site may help to raise accuracy [53], [22].

However, at this point, methodological challenges arise. First, a well-suited method for describing spatial context needs to be employed. Second, a semantically sound method for fusing spectral and context information needs to be employed. Third, computing context information raises the number of features used in classification. Therefore, many traditional classification approaches that are frequently used in ecological studies—such as maximum-likelihood classification (MLC)—tend to fail in such application scenarios [38]. Modern techniques from pattern recognition are required to tackle the methodological challenges and will be used herein. The primary objective of this study is to design a framework for producing highly accurate land-use maps in situations, where traditional multispectral data have to be used.

Manuscript received July 02, 2013; revised September 23, 2013; accepted October 29, 2013. Date of current version March 14, 2014.

A. C. Braun, H.-P. Bähr, S. Hinz, and U. Weidner are with the Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany (e-mail: andreas.ch.braun@kit.edu; hans-peter.baehr@kit.edu; stefan.hinz@kit.edu; uwe.weidner@kit.edu).

C. Rojas is with the Departamento de Geografia, Facultad de Arquitectura, Urbanismo y Geografia, University of Concepcion, Casilla 160-C, Correo 3, Concepción, Chile (e-mail: crojasq@udec.cl).

C. Echeverria is with the Landscape Ecology Laboratory, Facultad de Ciencias Forestales, University of Concepcion, Casilla 160-C, Correo 3, Concepción, Chile (e-mail: cristian.echeverria@udec.cl).

F. Rottensteiner, S. Kosov, and J. Niemeyer are with the Institute of Photogrammetrie and GeoInformation (IPI), Leibniz University of Hannover, 30167 Hannover, Germany (e-mail: rottensteiner@ipi.uni-hannover.de).

M. A. Arias is with the Environmental Sciences EULA-Chile Center, Environmental Science Faculty, University of Concepcion, Casilla 160-C, Correo 3, Concepción, Chile (e-mail: maaguayo@udec.cl).

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/JSTARS.2013.2293421

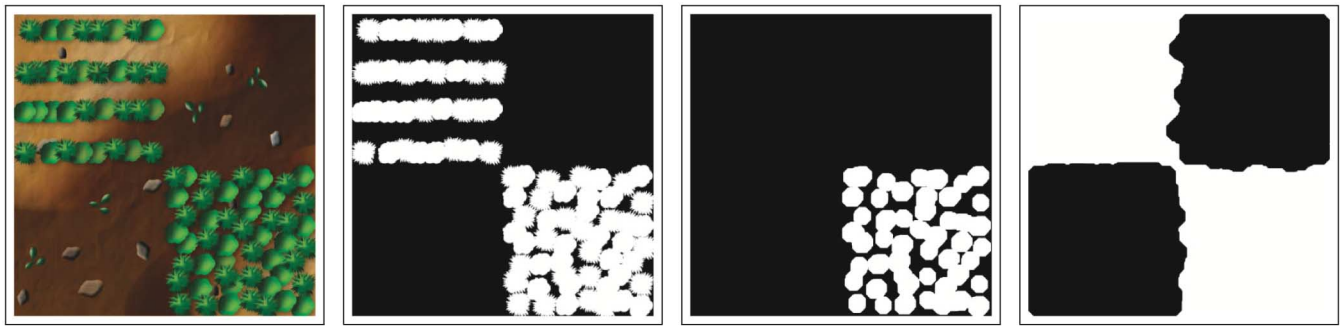

Fig. 1. Toy example: rationale for using the mathematical morphology. Plantation trees cannot be distinguished from the forest trees in the original image. Opening provides information to distinguish both. Open soils in plantations cannot be distinguished from the open soils elsewhere. Closing provides information to distinguish the open soils: (a) original image, (b) binary image, (c) opening, and (d) closing.

Design involves the combining state of the art methods from the pattern recognition that are suitable to enhance accuracy when processing the data with limited spectral information content and moderate spatial resolution. Secondary objectives are to provide information on the performances of import vector machine (IVM) and the relevance vector machine (RVM) in comparison to the support vector machine (SVM). Furthermore, the report is—to the knowledge of the author—the first to provide insight on the combination of IVM and RVM with kernel composition for data fusion.

Since Chile represents a typical case for developing countries, where fast and intense land-use change processes have to be analyzed with a poor remote-sensing database, the framework designed herein can be considered a case study of how to overcome data limitations in other regions of the global south.

II. Rationale for Modelling the Landscape Context by Context Features

In central Chile, native southern beech and sclerophyllous forests are replaced by plantations of *Pinus* and *Eucalyptus* trees. Fig. 1 shows a toy example to illustrate how context information can help to enhance separability. In the upper left corner, an exotic tree plantation is situated and in the lower right corner, a native forest can be seen. Due to the their high spectral similarity, trees in the upper left and lower right corner cannot be distinguished based on spectral information. A human interpreter uses the context to distinguish both, as plantations show a regular planting pattern, whereas native forests do not. The same accounts for the open soil in plantations in the upper left corner and the open soil in agriculture in the lower left corner. Fig. 1(b) shows a binary image produced on some threshold of the color image. In order to describe the spatial context into which pixels are integrated, where image openings and closings are helpful [44]. Fig. 1(c) shows an image opening produced with a structuring element SE (disk, size 10). As one can see, this feature clearly allows to distinguish trees in the upper left corner (which are background pixels in the binary image) from trees in the lower right corner (foreground pixels in the binary image). Fig. 1(d) shows an image closing produced with another SE (disk, size 25). This feature is not useful to distinguish the different kinds of trees. However, open soils between the plantation trees are clearly distinguished from open soils in the upper right and lower left corner. Hence, this feature helps to distinguish open soils in the plantations from, e.g., recently harvested areas. As can be seen, the size of SE determines its capacity to facilitate a particular separation problem. Since it is difficult to decide beforehand which sizes will be helpful, it should be attempted to use a whole range of differently sized SE.

For both the separation problems, a feature is computed which raises their separability through morphological operations. It can be concluded that for both cases, a reliable separation based on spectral features is not granted since the classes of interest show very similar spectral characteristics of individual pixels. The context in which pixels are embedded within the landscape, however, creates differences in the texture that are potentially highly significant for separation. Hence, a method should be employed to incorporate a description of the local neighborhoods of pixels into classification.

III. Methods

In the following, the methods used will be outlined. The emphasis is put on outlining the objectives for including a particular method into the entire framework. Mathematical foundations of particular methods can be found in the original publications and will be outlined briefly herein.

A. Extended Morphological Profiles

As pointed out in Section II, information on the landscape context represented by morphological context features can help to distinguish land-use classes if context features based on a range of various sizes are computed. Morphological profiles (MPs) calculate new features by grayscale morphology using openings and closings by reconstruction. From an image channel I of the 8-bit Landsat data, a stack of binary images are produced. This is achieved by using each integer in $0, \ldots, 255$ as a threshold. In each binary image of this stack, objects appear as hierarchically ordered and connected as components. An object bright enough to pass the 120 threshold will also be a connected

component of the 100 threshold. On this stack of images, image openings and closings by reconstruction are computed using SEs. SE are scaled by the parameter ζ_{SE}. Each connected component is tested against SE with given size ζ_{SE} using four different criteria.

1) *Area*: Tests whether the area of the connected component is larger than the area of SE.
2) *Diagonal*: Tests whether the diagonal bounding box of the connected component is larger than the diagonal of the bounding box of SE.
3) *Standard deviation*: Tests whether the standard deviation of the gray-level values of the pixels in the connected component is larger than ζ_{SE}.
4) *Inertia*: Tests whether the elongation of the connected component is larger than the elongation of SE according to the first moment invariant [28].

If a connected component passes the test—e.g., if the area of the connected component is larger than the area of SE—it is entirely preserved. If it fails, it is entirely integrated into the next brighter threshold level (closing by reconstruction) or the next darker threshold level (opening by reconstruction) [18]. Afterward, a feature image is constructed by reconverting the altered stack of binary images into a new grayscale image. This procedure can be applied to any integer data and is not limited to the 8-bit quantization used herein. The resulting morphological context features in the MP thus represent grayscale features. As shown in Section II, it is difficult to predict beforehand which size of SE will be helpful to separate classes. Therefore, a whole set of ζ_{SE} can be used to compute closings and openings by reconstruction. The resulting features provide a hierarchy of the spatial context. With increasing ζ_{SE}, larger areas on the ground will be unified in the image openings and closings. An area on the ground which is unified at a small ζ_{SE} will also be unified in larger ζ_{SE} in the hierarchy. An MP of s different ζ_{SE} of image channel I is then

$$\mathrm{MP} = \left[\mathrm{OP}_{\zeta_{\mathrm{SE}}s}, \mathrm{OP}_{\zeta_{\mathrm{SE}}s-1}, \ldots, I, \ldots, \mathrm{CL}_{\zeta_{\mathrm{SE}}s-1}, \mathrm{CL}_{\zeta_{\mathrm{SE}}s}\right]. \quad (1)$$

The extended morphological profile (EMP) of an image with m channels is the concatenation of the MPs of each image channel $\mathrm{EMP} = [\mathrm{MP}_1, \mathrm{MP}_2, \ldots, \mathrm{MP}_m]$. The number of additional features in the EMP is $m * s * 2 * 4$, since for m channels, s SE provide one closing and one opening by reconstruction, and four criteria are used at each size.[1] For the eight-channel Landsat data, 320 features are provided since five different ζ_{SE} are used ($\zeta_{\mathrm{SE}} \in [10^1, 10^2, 10^3, 10^4, 10^5]$ for area and diagonal, $\zeta_{\mathrm{SE}} \in [1.0, 1.5, 2.0, 3.0, 5.0]$ for standard deviation and inertia).

The MPs are the multiscale representations of the context pixels that are integrated. Although they have been originally invented for very-high-resolution (VHR) data, the application schemes they have been applied to are not limited to VHR data. The MPs and EMPs have been applied, e.g., by [3], [15] to very-high-resolution data, [2] and [22] on hyperspectral data, and combined optical LiDAR datasets by [39].

Numerous alternatives for computing context information have been published. Important examples are local gradient techniques [48], techniques based on wavelets, anisotropic partial differential equations [52], or multiscale features such as empirical mode decomposition which are most similar to the EMP features used herein [19]. Another possibility of considering pixels context in classification is rule-based object-oriented classification [4]. Graphical models [32], [16] are another technique of modeling context in classification (cf. Section III-F).

However, many of these alternatives have been applied to a single type of dataset only. These datasets are frequently modern remote-sensing datasets such as hyperspectral data. Especially for moderate resolution data, very few studies have been published. Thus, the applicability of these alternatives to Landsat data is less clear than for EMPs which have successfully been applied to almost any kind of remote-sensing dataset. This gives reason to assume that they are generally applicable to Landsat data as well.

B. Feature Selection

As stated above, the EMP consists of 320 additional features. As shown for a toy example, some may provide highly discriminative features for certain classes while others may not be helpful (cf. Section II). Therefore, it should be aimed to remove some nonsignificant features by feature extraction or feature selection techniques [33]. The EMP is a hierarchical representation of spatial context in the image channels. Applying feature extraction techniques based on data rotations disrupts this hierarchy, since more than one hierarchical level is projected onto the same rotation component. Thus, instead of using rotation-based techniques such as PCA, ICA, or kernel-PCA [13], more straightforward selection techniques seem indicated since they maintain the hierarchy. Once image classes are assigned and represented by training areas in the image, a forward selection technique is applied [55]. For this purpose, the Jeffries-Matusita δ_{JMD} is used which describes the a priori separability of classes.[2] In thefirst step, for each pair of 2 out of C classes, the δ_{JMD} based only on the spectral features is computed and written into a $C \times C$ matrix M^{+0}_{spec}. Then, a single feature of the EMP is added to the dataset and another matrix M^{+1}_{spec} is computed. The feature is used only if $\Sigma M^{+1}_{\mathrm{spec}} - \Sigma M^{+0}_{\mathrm{spec}}$ is positive. In this case, this EMP feature produces a net improvement of the separability δ_{JMD} summed over all classes. This test is performed for each feature and the EMP is reduced to the features that fulfill the criterion.

Forward selection techniques are used in the remote-sensing community by [43] to select hyperspectral channels and by [47] to select classifiers for a multiple classifier system (MCS).

C. Data Representation and Fusion by Kernels

After reducing the EMP to a set of presumably helpful features, the features can be used to construct a feature space in which they can be classified. At the same time, both information types—the spectral channels and the spatial EMP features—need to be fused. A straightforward approach for data fusion

[1]Note that this number excludes the spectral image channel I itself which are included in (1) and only counts the *additional* features of the EMP.

[2]Note that δ_{JMD} assumes that classes are Gaussian distributed. The validity of this assumption was evaluated by interpreting the scatter-plots of the main classes.

would be to simply join both the information types in a large data matrix (concatenation CC). However, since the spectral and spatial information are semantically different, this method may not be optimal. First, the EMP features should not be treated as if they were merely some extra channels, whereas in fact, they are a different kind of information. Second, only one distance measure will be computed on the concatenated feature matrix. The operator has no option to define, whether he considers the distance in the spectral data or the distance in the EMP to be more relevant for classification [7]. For instance, a pronounced difference between two classes in a single spectral channel may be disguised by random differences in a series of EMP features—or vice versa. An alternative is given using kernels. The kernels represent a distance measure between the two input feature vectors x_i and x_j after implicit transformation $\phi(x_i)$ to a high-dimensional feature space

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) = \exp(-\sigma \|x_i - x_j\|^2). \quad (2)$$

Although the number of features in x_i, x_j can be large (e.g., 320 features in the EMP), the resulting kernel value $K(x_i, x_j)$ is a single scalar [42]. Herein, the radial basis function (RBF) kernel is used (2). The RBF depends on a free parameter σ that is tuned by cross-validation grid search in training ($\sigma \in [2^{-15}, \ldots, 2^5]$). The implicit transformation[3] $\phi(x_i)$ has the advantage of making difficult classification problems more easily separable [11]. After embedding the data into such a feature space induced by kernels—called reproducing kernel hilbert space (RKHS)—data fusion is easily tackled. The theory of Kernels and RKHS states that kernels can be combined (e.g., by addition or multiplication) to form another valid kernel [12]. Kernel composition (KCOMP) exploits this fact for data fusion

$$\text{DSUM}: \quad K_{\text{CMP}} = K_{\text{SPC}} + K_{\text{EMP}} \quad (3)$$

$$\text{WSUM}: \quad K_{\text{CMP}} = f_1 \cdot K_{\text{SPC}} + f_2 \cdot K_{\text{EMP}} \quad (4)$$

$$\text{PROD}: \quad K_{\text{CMP}} = K_{\text{SPC}} \times K_{\text{EMP}}. \quad (5)$$

For concatenation, only one RBF kernel can be used and σ is tuned jointly over the spectral data and the EMP. However, a different σ may be required for the spectral than for the EMP data. In kernel composition, one RBF kernel K_{SPC} can be computed on the spectral data and another RBF kernel K_{EMP} on the EMP. For both kernels, σ is tuned independently (σ_{SPC} and σ_{EMP}). Then, they are combined to a composite kernel K_{CMP} by direct summation (DSUM), weighted summation (WSUM), or products (PROD).

Kernel composition thus allows a semantically sound data fusion. Spectral and spatial data are used to separately define two different distance measures represented by the kernels K_{SPC} and K_{EMP}. Then, the two distance measures are combined. Differences in one data source are thus not concealed by noise in the other. The data sources are therefore treated differently since each of them are represented by a particular kernel and the operator can control their combination. This is especially true for WSUM, where the operator can define weights f_1, f_2 to indicate the relevance of the respective distance measure for classification. However, since no a priori information on the relevance of the spectral data in comparison to spatial data is available, these weights are tuned by grid search. After tuning the kernel parameter σ, f_1, f_2 are determined by cross validation evaluating the performance of $f_1 \in [0.05, 0.10, \ldots, 0.95]$ and $f_2 = 1 - f_1$.

Composed kernels have been used by [12] for hyperspectral–spatial classification, by [9] for classification of fused hyperspectral and LiDAR datasets, and by [34] for using backscatter information of SAR data jointly with local neighborhoods.

D. Classification

Having represented the data in RKHS using composed kernels, appropriate classification is the next task. Since traditional classifiers are not suitable for RKHS, kernel-based classification has to be employed. The framework designed herein is therefore based on three kernel-based classifiers—the support vector machine (SVM) [5], [6], the import vector machine (IVM) [57], and the relevance vector machine (RVM) [49]. The three methods all follow a similar principle. Given a training dataset X with n training vectors x_i and scalar labels $y_i,\ i \in [1, \ldots, n]$, all three models build a nonlinear feature space based on the K_{CMP}. Within this feature space, a separating plane is optimized to assign unknown data x_k to the classes y_k. Since training data in the ecological applications are frequently scarce, the separating plan will depend only on a small subset $V \subset X$ of training samples. For this purpose, each training data point is assigned a linear weight ω_i that controls to which extent the training point influences the separating plane. In order to optimize the separating plane and to make it dependent only on a few data points, ω_i is defined to be positive only for the subset V. For the remaining points, $\omega_i = 0$. ω_0 is a bias parameter which is needed to project the separating plane from the origin to the data free space between the two classes

$$y_k = f(x_k; \omega_i) = \sum_{i=1}^{n} \omega_i K(x_k, x_i) + \omega_0. \quad (6)$$

Although they are based on the same model, SVM, IVM, and RVM do have some important differences. The main difference, which is relevant herein, is the choice of subset V—i.e., for which type of points $\omega_i > 0$. The SVM uses the points closest to the other class (the borderline points). The IVM uses some points from the entire distribution, and the RVM uses a few prototype points to establish the separating plane. Since the separating planes are differently established, each classifier may be correct on some samples on which both other classifiers fail. For a thorough discussion of the classifiers differences, cf. [8], [29], and for mathematical foundations see the original publications [5], [57], [49]. Their common formulation makes the classifiers well combinable. For the major part of the feature space, the separating planes as defined by the three methods take identical decisions on the class membership. However, the three methods define their separating planes on different types of input points

[3]Readers who are not familiar with kernel-based classification are recommended to watch the Youtube Video *SVM with polynomial kernel visualization* which perfectly illustrates the concept, http://www.youtube.com/watch?v=3liCbRZPrZA.

(cf. Figs. 1 and 4 in [8]). Therefore, for a minor part of the feature space, different decisions are taken. For this reason, for a few points, some classifiers take correct decisions, whereas others do not. Thus, combining the outputs of the three classifiers will correct the decisions on points that are falsely classified by one method but correctly classified by the other two methods. Hence, the different formulation of the separating plane by the three methods gives reason to expect a combination of these classifier to perform better than each classifier on its own.

Numerous theoretical and applied studies have been published on kernel-based classification and especially on SVM. For an overview of SVM in remote sensing, [35] is recommended and some recent developments are found in [40]. A theoretical overview of kernel-based classification in remote sensing is given in [11]. Thorough comparisons of SVM, IVM, and RVM are given in [8] and [29].

E. Multiple Classifier System

Three different kernel-based classifiers are employed. It should be expected that each of them performs better by separating some classes and worse by separating others [56]. Thus, MCSs aim to improve classification accuracy after classification by applying a fusion rule which combines the decision values (i.e., votes) of the base classifiers (SVM, IVM, and RVM) [1]. Majority voting (MV) is one example of a fusion rule for a MCS which assigns each pixel a new class label which corresponds to the label most frequently assigned by the base classifiers. Fuzzy MV (FMV) now fuzzifies MV in order to improve robustness [41]. For each pixel, the posteriors $p_b(y_c|x)$ of each of the $b = [1, 2, 3]$ base classifier are rescaled to fuzzy membership functions q_b. Two thresholds s_1 and s_2 are defined and $q_b = 0$ if $p_b < s_1$, $q_b = 1$ if $p_b > s_2$, and else $q_b = (p_b - s_1)/(s_2 - s_1)$. Then, the fuzzy membership functions q_b of the classifiers are ranked by $j_b \in [3, 2, 1]$. j_b represent the order of the bth classifier after ranking the fuzzy membership functions q_b of the pixel, for SVM, IVM, and RVM in a descending order. The fuzzy membership functions q_b are then converted into weights by

$$w_b = q_b \cdot \left(\frac{j_b}{3}\right) - q_b \cdot \left(\frac{j_b - 1}{3}\right). \tag{7}$$

New posteriors are computed according to sum of the weights of the three classifiers $p_{\mathrm{fmv}}(y_c|x) = \arg\max(\Sigma w_b p_b)$. Thus, s_1 truncates the left tail of the posteriors probability density function. Posteriors smaller than s_1 are considered irrelevantly small and will be entirely ignored in the MCS[4]. s_2 similarly evaluates the right tail. Posteriors larger than s_2 are considered as fully relevant and thus set to 1. Herein, $s_1 = 0.1$ and $s_2 = 0.5$ are used. Each classifier influences this function not only based on the label it is assigned to the pixel, but also on its posterior probability. According to these new posteriors, a combined class label is assigned. Thus, FMV does not strictly decide on the basis of the majority vote, neglecting the vote of the minority. Instead, it permits each classifier to influence the combined class label according to its posterior probability $p_b(y_c|x)$. These posterior probabilities are fuzzified in order to compute a joint new posterior $p_{\mathrm{fmv}}(y_c|x)$. Thus, the minority may influence the decision more strongly, if its confidence is very high and the majorities confidence is low. Therefore, a superior performance than for traditional majority voting is assumed.

MCSs are subject to intensive research in remote sensing. An overview on theoretical and applied aspects is given in [1] and [10]. Fuzzy majority techniques are used by [36] to unify the decisions of a spectral spatial classification approach and by [51] to unify statistical and neural classifiers in building extraction.

F. Conditional Random Fields

So far, the spatial context into which pixels are integrated is exploited implicitly through the EMP. Although this implicit representation is supposed to raise accuracy [2], [22], the approach has one drawback. The assignment of a class value c to the pixel i is independent from class values of neighboring pixels. Hence, further accuracy improvement can be achieved by considering a neighborhood around individual pixels classifying each pixel in connectivity with its neighbor [31]. This approach will support the classification in case when the EMP classifier provides near equal potentials for pixel i to have both classes $c = A$ and $c = B$. In this case, the label will be chosen based on the labels of neighboring pixels. Conditional random field (CRF) is a standard technique to model dependencies between the random variables in stochastic problems [32]. It makes use of a graphical representation of the data $G(N, E)$, where nodes N represent random variables, corresponding to each pixel; whereas edges E, which connect the nodes, represent interaction between these random variables. The joint posterior $p(y|x)$ can be modelled through CRF as follows:

$$p(y|x) = \frac{1}{Z}\exp\left(\sum_{i\in N} A_i(y_i, x) + \sum_{i,j\in E} I_{ij}(y_i, y_j, x)\right). \tag{8}$$

In (8), $A_i(y_i, x)$ are the association potentials linking the observations to the class label at pixel i. $I_{ij}(y_i, y_j, x)$ are the interaction potentials modeling the dependencies between the class labels at two connected pixels i and j and the data x. Z is a normalizing constant, also called the partition function. On the set of all edges E for our task, we define the eight nearest neighborhood systems and thus

$$\sum_{i,j\in E} I_{ij}(y_i, y_j, x) \equiv \sum_{i\in N}\sum_{j\in \mathcal{N}_i} I_{ij}(y_i, y_j, x). \tag{9}$$

The association potentials $A_i(y_i, x)$ are related to the probability of a label y_i taking a value c given the data x by $A_i(y_i, x) = \log p(y_i = c|f_i(x))$ [31], where the image data are represented by pixel-wise feature vectors $f_i(x)$ that may depend on all the observations x. Note that both the definition of the features and the dimension of the feature vectors $f_i(x)$ may vary with the dataset. We use posteriors of the MCS directly for the association potentials. The interaction potentials $I_{ij}(y_i, y_j, x)$ describe how likely the pair of neighboring sites i and j is to take the labels $(x_i; x_j) = (c; c')$ given the data: $I_{ij}(y_i, y_j, x) = \log p(x_i = c; x_j = c'|y)$ [31]. Our data-dependant definition

[4] s_1 and s_2 can be arbitrarily chosen between $[0, \ldots, 1]$ as long as $s_1 \geq s_2$ and have been set $s_1 = 0.1$ and $s_2 = 0.7$ herein.

of $I_{ij}(y_i, y_j, x) \equiv I_{ij}(y_i, y_j, d_{ij})$ is obtained by applying a penalization depending on the Euclidean distance $d_{i,j} = \|f_i(x) - f_j(x)\|$ of the node feature vectors f_i and f_j.

CRFs have been successfully applied to remotely sensed data by [30], [27], [54], and [26].

G. Entire Framework

The framework is composed of five phases in which the complexity of the methods used is subsequently increased.

1) *Phase 1*: Basic classification using the spectral information.
2) *Phase 2*: Classification using the spectral information and EMP.
3) *Phase 3*: Classification using the spectral information and EMP applying feature selection.
4) *Phase 4*: Application of the MCS to the results of SVM, IVM, and RVM.
5) *Phase 5*: Application of the CRF to the results of MCS.

It is aimed to maximize discriminative potential in terms of overall accuracy. Thus, the frameworks quality will be assessed at each stage.

H. Quality Evaluation

Before applying the developed framework, each subset has been manually classified by a human operator (the results of the manual classification can be seen in Fig. 2). Due to the computational expense, not all labelled pixels are used in training. From the manually classified image, a subset of 300 pixels were randomly selected for training and another disjoint subset of 300 pixels was randomly selected for quality evaluation. To compare the quality of the results, confusion matrices on the basis of the control data are computed. There are various definitions of quality in remote sensing and thus various indices [23], [24]. Since [45] postulate that results in ecological applications to be at least 90% *accurate*, quality is assessed via the overall accuracy (OAA). For the design of the framework, it is further interesting to see the improvement in the OAA figure between the two subsequent phases. For this purpose, the increase in OAA is assessed through $\Delta_{p2/1} = \mathrm{OAA}_{p2} - \mathrm{OAA}_{p1}$, where OAA_{p1} and OAA_{p2} are the best results yielded on a subset in each phase.

IV. Results

Entire Landsat scenes are too large to be processed within a convenient time when designing the framework. Hence, three subsets are used (for classes and coloration, cf. Table I). The first subset Su1 [cf. Fig. 2(a)] is chosen to focus on a mosaic of native forest classes and exotic tree plantations at different stages. The second subset Su2 [cf. Fig. 2(b)] covers the suburban, urban classes, and classes near the shoreline. The third subset Su3 [cf. Fig. 2(c)] focuses on areas where forests, plantations, and agricultural areas meet. All subsets are of size 383×430 pixels. It is assumed that these three subsets adequately cover the complexity of the entire scenario. Quality evaluation is performed as outlined in Section III-H and summarized in Table II.

A. Phase 1: Basic Classification Using Spectral Information

At first, spectral information is classified without the EMP. The results are given in Fig. 2(g)–(i). Visual results show a rough impression since they are strongly affected by salt-and-pepper noise, which is most pronounced in Su1 and Su3. In Su3, the agricultural areas in the south-center disappear almost entirely between the plantation and agriculture classes. In the Su1, the harvested area is mixed up with the plantation classes. In Su2, native forests in the south are very much mixed up with plantations which surround them. For each subset, only unsatisfactory OAA values could be yielded (cf. Table II). However, traditional classifiers such as MLC perform worse than modern classifiers. Although hardly used in the pattern recognition community, the MLC is still employed in some applied remote-sensing studies [20]. The $\Delta_{p1/0}$ values show that modern pattern recognition techniques yield significant advantages.

B. Phase 2: Classification Using Spectral Information and EMP

Next, the EMP is amended to the dataset using the kernel composition. No feature extraction was performed before classification. The classification results of SVM are given in Fig. 2(j)–(l). In contrast to the results presented in Section IV-A, the classification results are much smoother and less affected by salt-and-pepper noise. This is most obvious for Su1, e.g., in the large plantation (type 3) in the center or in the south, in the harvested area in the west, and the forested areas in the southeast. It should be noted that besides smoothing (in the sense of removing single pixels of one class within large areas of other classes), the areas seem to be better ordered. Large areas of a certain class are much more coherent than in the previous result. This can be observed most clearly for the agricultural classes in the south-center of Su3 and the forest classes in Su3. The same accounts for the large forest areas in the center of Su1 or the area below the city in Su2. As the $\Delta_{p2/1}$ values in Table II point out, the EMP thus significantly improves the performance of the framework.

C. Phase 3: Classification Using Spectral Information and EMP Applying Feature Selection

We applied feature extraction methods described in Section III-B. For Su1, 208 features from the EMP were selected; for Su2, 180 were selected; and for Su3, 188 features were selected. Visual results are not shown since the performance gains are only slight for two of three results. As the $\Delta_{p3/2}$ values in Table II make clear, the feature selection improved the results significantly on Su3 but not on Su1 and Su2. However, giving explanations for this improvement on the basis of three results is not reasonable.

D. Phase 4: Application of the MCS to the Results of SVM, IVM, and RVM

In the next phase, the decision outputs were fused with the MCS based on the FMV. We fused the best results based on kernel composition of each of the three classifiers. Again, visual results are not shown since performance gains are hardly visible in the images. According to the $\Delta_{p4/3}$, the MCS thus brought a

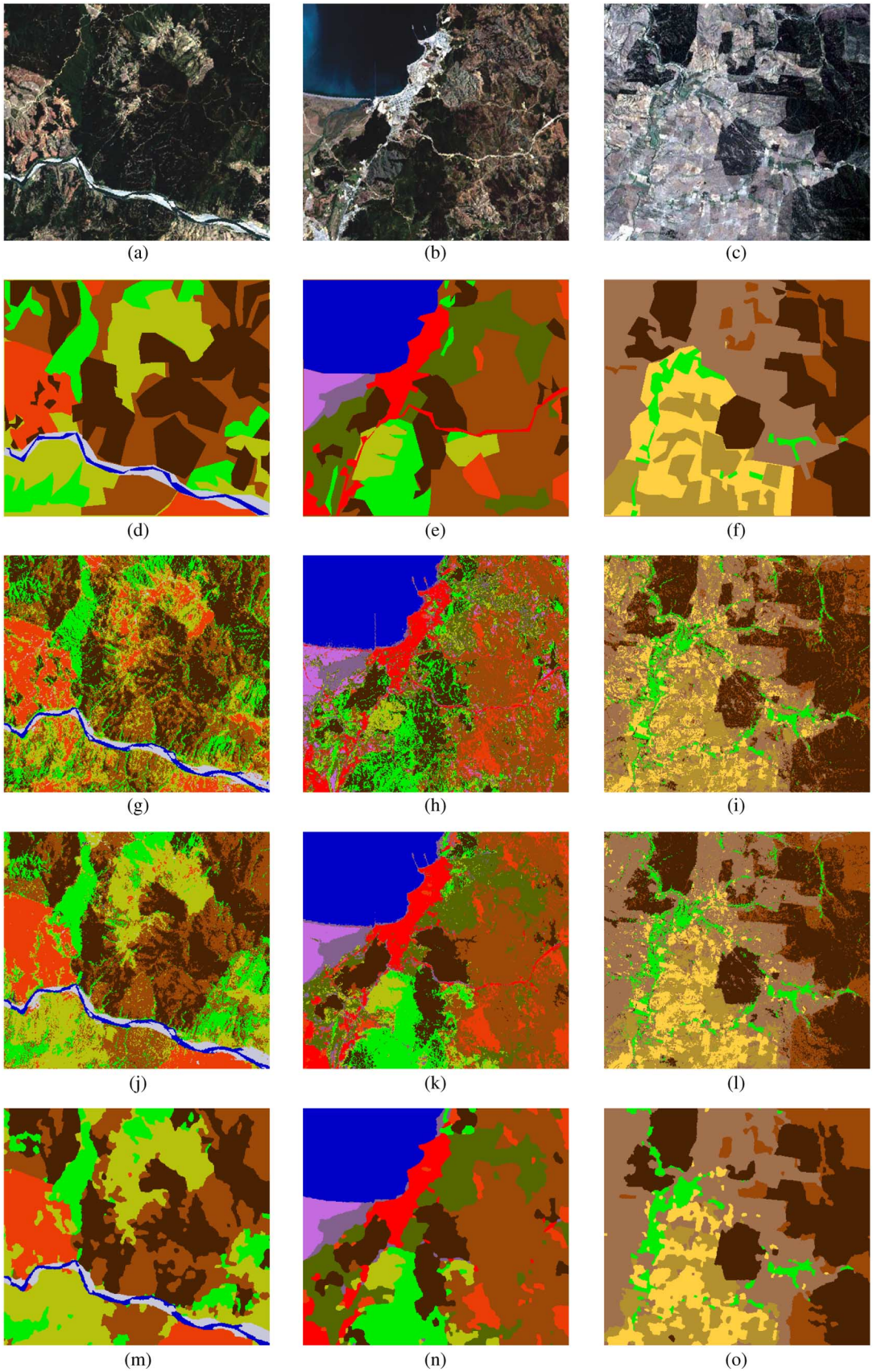

Fig. 2. Left: forestry subset (Su1), Center: semi-urban subset (Su2), and Right: agricultural subset (Su3). First row: truecolor visualization. Second row: result of manual classification (reference). Third row: initial classification (SVM on spectral data only), Fourth row: classification after adding EMP (SVM), and Fifth row: final classification (entire framework).

TABLE I
LEGEND FOR THE CLASSIFICATION RESULTS SHOWN IN FIG. 2

Color	Name	SU1	SU2	SU3
	Forests	✓	✓	✓
	Plantations type 1	✓	✓	✓
	Plantations type 2	✓	✓	✓
	Plantations type 3	✓	✓	×
	Clear-cut	✓	✓	×
	Water	✓	✓	×
	Sediment	✓	×	×
	Settlement	×	✓	×
	Wetlands type 1	×	✓	×
	Wetlands type 2	×	✓	×
	Ecotone (forest/plant.)	×	✓	×
	Agriculture type 1	×	×	✓
	Agriculture type 2	×	×	✓
	Open soil	×	×	✓

✓ indicates that a class is defined in subset Su1, Su2 or Su3 and × indicates that this class is not defined in the respective subset.

TABLE II
OVERALL ACCURACY (OAA) VALUES DURING THE FRAMEWORKS PHASES (P)

P	Appr.	Data	F.E.	D.F.	OAA		
					SU1	SU2	SU3
0	MLC	SC	—	—	58.5	65.4	51.9
1	SVM	SC	—	—	**61.9**	**78.5**	71.0
1	IVM	SC	—	—	61.2	78.3	71.6
1	RVM	SC	—	—	61.0	77.7	**71.8**
1	$\Delta_{p1/0}$				3.4	13.1	19.9
2	SVM	SC+SP	—	CC	76.2	93.1	83.2
2	IVM	SC+SP	—	CC	80.0	90.1	70.2
2	RVM	SC+SP	—	CC	75.1	84.2	70.3
2	SVM	SC+SP	—	KC	78.3(D)	95.4(W)	86.3(W)
2	IVM	SC+SP	—	KC	**81.7(P)**	**95.7(D)**	77.3(D)
2	RVM	SC+SP	—	KC	77.8(P)	84.3(D)	**88.5(P)**
2	$\Delta_{p2/1}$				19.8	17.2	16.7
3	SVM	SC+SP	PCA	KC	76.4(D)	93.2(W)	90.2(W)
3	IVM	SC+SP	PCA	KC	77.2(D)	93.5(P)	92.1(D)
3	RVM	SC+SP	PCA	KC	76.7(W)	92.5(W)	91.4(D)
3	SVM	SC+SP	JMD	KC	**82.6(W)**	**95.8(W)**	91.2(W)
3	IVM	SC+SP	JMD	KC	81.6(W)	95.7(D)	**94.0(D)**
3	RVM	SC+SP	JMD	KC	78.3(W)	94.5(W)	93.1(W)
3	$\Delta_{p3/2}$				0.9	0.2	5.6
4	MCS	$p(y_c\|x)$	—	FMV	**83.3**	**96.0**	**96.0**
4	$\Delta_{p4/3}$				0.7	0.2	1.9
5	CRF	A_{ij}+I_{ij}	—	—	**87.9**	**97.5**	**98.4**
5	$\Delta_{p5/4}$				4.6	1.5	2.4

SC: spectral information, SP: spatial information (EMP), F.E.: feature extraction, D.F.: data fusion, CC: concatenation, KC: kernel composition (D: DSUM, W: WSUM, and P: PROD). Boldface figures indicate the result that causes an improvement to the previous phase $\Delta_{p/p-1}$ value.

slight improvement for all subsets. Again, the improvement on Su3 is most pronounced.

E. Phase 5: Application of the CRF to the Results of MCS

Next, the CRF is applied on the dataset. The posteriors of the MCS are used as the association potential A_{ij} of the CRF. Visual results are given in Fig. 2(m)–(o). The CRF produces a remarkable homogenization of the classification result. Salt-and-pepper noise is entirely suppressed. Furthermore, areas with identical land use are aggregated to large regions in the images. Although some details are lost—e.g., the roads or small assemblages of trees—these results best describe the land use at the landscape level. The benefits of the CRF can also be seen in the quantitative results. The CRF thus produces significant performance gains for

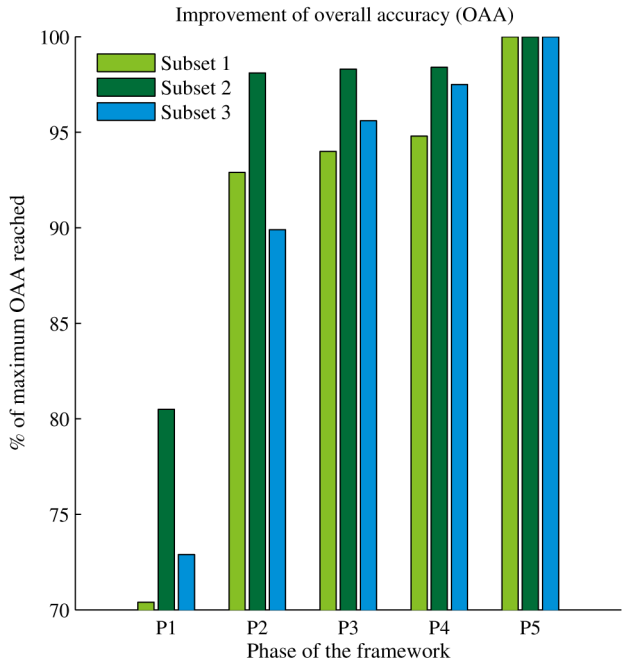

Fig. 3. Accuracy (OAA) yielded at each phase of the framework as percentage of the total OAA yielded for the respective subset. Vertical axis is scaled to [70, . . . ,100]%.

each subset. The most pronounced gain was observed for Su1—the subset with the lowest accuracy value after the fourth phase. Note that on the other subsets, similar gains as for Su1 are not possible after the fourth phase since the 100% limit would be exceeded.

V. DISCUSSION

Significant total OAA improvements—i.e., the $\Delta_{p5/1}$ values—are observed. The performance of the framework exceeds by far the performance of the MLC classifier which is still used in some studies [20]. For Su1, an improvement of $\Delta_{p5/1} = 26.0\%\text{pts}$ is observed. For Su2, the improvement is $\Delta_{p5/1} = 19.0\%\text{pts}$ and for Su3, $\Delta_{p5/1} = 26.6\%\text{pts}$. Thus, for all subsets, performance gains are at least close to 20%pts or even higher. All results are close to the 90% OAA claim of [45] or even surpass it. It is interesting to see at which phases performances are raised. The biggest gain for all subsets comes from using the spatial context. The second biggest gain comes from the application of the CRF. The two final phases of the framework (MCS and CRF) raise accuracy significantly beyond base classification. The $\Delta_{p5/3}$ values are $\Delta_{p5/1} = 5.3\%\text{pts}$ for Su1, $\Delta_{p5/1} = 1.7\%\text{pts}$ for Su2, and $\Delta_{p5/1} = 4.3\%\text{pts}$ for Su3. These circumstances are visualized in Fig. 3 which plots the accuracy gained at each phase of the framework as a percentage of the final accuracy yielded for that subset. During phase 3, only Su3 is significantly improved with respect to the accuracy yielded by phase 2 (EMP). However, the feature selection step in phase 3 improved the computation time. Training on the full EMP was prolonged on average by the factor 1.04 for SVM, 1.14 for IVM, and 1.10 for RVM. Evaluating which features performed best, few conclusions can be drawn. For each subset, every image channel, every

criterion, and every ζ_{SE} value contributed to the features selected by the forward selection. Thus, no features can be excluded a priori. Hence, the EMP needs to be fully computed with all features. Concerning the feature selection, it is shown that reducing redundant features improves accuracy. Only one result (IVM on Su1) yielded a higher accuracy (0.1%pts.) with the full EMP than with the JMD selected EMP. As expected, selecting features on the basis of JMD is superior to PCA. All JMD results are superior to PCA. Furthermore, the PCA is superior to using the full profile for only four results (all results on Su3 and the RVM on Su2). Thus, the rotating features by PCA deteriorates the accuracy for most cases—confirming the conclusions of [14]. There is no clear tendency with respcect to the best classifiers. The SVM, IVM, and RVM all contribute to the relatively best results at each phase (boldface letters in Table II). However, not all results that have been produced during the algorithm development are shown. Relevant are the results at phase 3—the final base classification phase. We identified the three best results during this phase. For Su1, it is SVM-WSUM, IVM-WSUM, and IVM-PROD. For Su2, it is SVM-WSUM, IVM-DSUM, and IVM-PROD and for Su3, it is IVM-DSUM, IVM-WSUM, and IVM-PROD. Thus, the RVM is never represented in the best results of phase 3 and all but two of the best results are yielded by IVM. Similar results are found with respect to the best kernel composition type. DSUM, WSUM, and PROD are all represented in the best results at each phase. However, for the best results during the base classification phase, most results are yielded by WSUM and two by DSUM. The PROD contributes three results which all yield the third rank. The optimal weight parameters chosen for WSUM varied between $f_1 = 0.4$ and $f_1 = 0.7$. For the framework, three different classifiers (SVM, IVM, and RVM) and three different kernel composition approaches for data fusion (DSUM, WSUM, and PROD) are evaluated. Therefore, nine results are available for each subset. In total, the difference between the accuracy values yielded are rather small. The standard deviation of OAA σ_{OAA} is $\sigma_{\mathrm{OAA}} = 2.0$ for Su1, $\sigma_{\mathrm{OAA}} = 2.9$ for Su2, and $\sigma_{\mathrm{OAA}} = 2.6$ for Su3. It is further interesting to see which classifiers performed best. The SVM–WSUM approach yielded the highest accuracy values for each subset (cf. Table II). Furthermore, it is assessed that kernel composition type performed best on each classifier. Frequently, the best fusion type is WSUM. Only two exceptions are observed. For Su1, the PROD performed better than WSUM with IVM and RVM (cf. Table II). Concerning the data fusion type, kernel composition is superior to concatenation. No concatenation result outperformed the respective kernel composition result. The σ values for K_{SPC} and K_{EMP} account for this finding. $\sigma_{\mathrm{SPC}} = \sigma_{\mathrm{EMP}}$ holds for only two of nine results in Phase 3 (Su1-IVM and Su1-RVM). For the seven other results, both σ differed significantly (e.g., $\sigma_{\mathrm{SPC}} = 2^5$ and $\sigma_{\mathrm{SPC}} = 2^{-8}$ for Su3-SVM). Hence, tuning a particular kernel on each data source improves the results.

VI. Conclusion and Outlook

A framework for yielding highly accurate land-use maps in moderate spectral and spatial resolution data is proposed. The accuracy values on three subsets of an entire deforestation monitoring scenario in Chile are improved between 19.9%pts and 26.6%pts. Major improvements are achieved by computing context features based on mathematical morphology. Further improvements are achieved by combining three base classifiers (SVM, IVM, RVM) to a new decision using an MCS and postprocessing the result using a CRF. No definite decision can be made which classifier is best. However, SVM and IVM tend to perform better than RVM. It can be concluded that the accuracy gains are achieved by additional features and combining and postprocessing the classifiers results. The framework is ready to be applied for land-use change assessments and deforestation monitoring scenarios.

References

[1] J. A. Benediktsson, J. Chanussot, and M. Fauvel, "Multiple classifier systems in remote sensing: From basics to recent developments," in in *Multiple Classifier Systems*, M. Haindl, J. Kittler, and F. Roli, Eds., New York, NY, USA: Springer, 2007, pp. 501–512.

[2] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[3] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[4] T. Blaschke, S. Lang, E. Lorup, J. Strobl, and P. Zeil, "Object-oriented image processing in an integrated GIS-remote sensing environment and perspectives for environmental applications," *Environ. Inf. Plann. Polit. Public*, vol. 2, pp. 555–570, 2000.

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

[6] A. C. Braun, U. Weidner, and S. Hinz, "Support vector machines for vegetation classification—A revision," *Photogramm. Fernerkundung Geoinf.*, vol. 2010, no. 4, pp. 273–281, 2010.

[7] A. C. Braun, U. Weidner, and S. Hinz, "Classifying roof materials using data fusion through Kernel composition comparing v-SVM, and one-class SVM," in *Proc. IEEE Joint Urban Remote Sens. Event, JURSE*, Munich, Germany, 2011.

[8] A. C. Braun, U. Weidner, and S. Hinz, "Classification in high-dimensional feature spaces—Assessment using SVM, IVM and RVM with focus on simulated EnMAP data," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 436–443, Apr. 2012.

[9] A. C. Braun, U. Weidner, B. Jutzi, and S. Hinz, "Kernel composition with the one-against-one cascade for integrating external knowledge into SVM classification," *Photogramm. Fernerkundung Geoinf.*, vol. 2012, no. 4, pp. 371–384, 2012.

[10] G. J. Briem, J. A. Benediktsson, and J. R. Sveinsson, "Multiple classifiers applied to multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2291–2299, Oct. 2002.

[11] G. Camps Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[12] G. Camps Valls, L. Gomez Chova, J. Muñoz Mari, J. Vila Frances, and J. Calpe Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[13] L. J. Cao, K. S. Chua, W. K. Chong, H. P. Lee, and Q. M. Gu, "A comparison of PCA, KPCA and ICA for dimensionality reduction in support vector machine," *Neurocomputing*, vol. 55, no. 1, pp. 321–336, 2003.

[14] T. Castaings, B. Waske, J. A. Benediktsson, and J. Chanussot, "On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5921–5939, 2010.

[15] J. Chanussot, J. A. Benediktsson, and M. Fauvel, "Classification of remote sensing images from urban areas using a fuzzy possibilistic model," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 40–44, Jan. 2006.

[16] R. Chellappa and S. Chatterjee, "Classification of textures using Gaussian Markov random fields," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 4, pp. 959–963, Aug. 1985.

[17] W. B. Cohen and S. N. Goward, "Landsat's role in ecological applications of remote sensing," *Bioscience*, vol. 54, no. 6, pp. 535–545, 2004.
[18] M. Dalla-Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[19] B. Demir and S. Ertürk, "Empirical mode decomposition of hyperspectral images for support vector machine classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4071–4084, Nov. 2010.
[20] E. R. DeRoeck, N. E. C. Verhoest, M. H. Miya, H. Lievens, O. Batelaan, A. Thomas, and L. Brendonck, "Remote sensing and wetland ecology: A South African case study," *Sensors*, vol. 8, no. 5, pp. 3542–3556, 2008.
[21] C. Echeverria, D. Coomes, J. Salas, J. M. Rey Benayas, A. Lara, and A. Newton, "Rapid deforestation and fragmentation of Chilean temperate forests," *Biol. Conservat.*, vol. 130, no. 4, pp. 481–494, 2006.
[22] 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. 15th IEEE Int. Conf. Image Process.* (ICIP 2008), 2008, pp. 1884–1887.
[23] G. M. Foody, "Status of land cover classification accuracy assessment," *Remote Sens. Environ.*, vol. 80, no. 1, pp. 185–201, 2002.
[24] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–634, 2004.
[25] M. Govender, K. Chetty, and H. Bulcock, "A review of hyperspectral remote sensing and its application in vegetation and water resource studies," *Water SA*, vol. 33, no. 2, pp. 145–152, 2007.
[26] W. J. He, M. Jäger, A. Reigber, and O. Hellwich, "Building extraction from polarimetric SAR data using mean shift, and conditional random fields," in *Proc. 7th Eur. Conf. Synth. Aperture Radar* (EUSAR), VDE, 2008, pp. 1–4.
[27] T. Hoberg, F. Rottensteiner, and C. Heipke, "Classification of multitemporal remote sensing data using conditional random fields," in *Proc. 2010 IEEE IAPR Workshop Pattern Recogn. Remote Sens.* (PRRS), 2010, pp. 1–4.
[28] M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. Inf. Theory*, vol. 8, no. 2, pp. 179–187, Feb. 1962.
[29] A. C. Karaca, A. Ertürk, M. K. Gullu, and S. Ertürk, "Comparative evaluation of vector machine based hyperspectral classification methods," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* (IGARSS), 2012, pp. 4970–4973.
[30] S. Kosov, F. Rottensteiner, and C. Heipke, "3D classification of crossroads from multiple aerial images using conditional random fields," in *Proc. IEEE IAPR Workshop Pattern Recogn. Remote Sens.* (PRRS), 2012, pp. 1–4.
[31] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vis.*, vol. 68, no. 2, pp. 179–201, 2006.
[32] J. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. 18th Int. Conf. Machine Learning*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289.
[33] D. D. Lewis, "Feature selection, and feature extraction for text categorization," in *Assoc. Comput. Ling. Proc. Workshop Speech Nat. Language,* 1992, pp. 212–217.
[34] 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.
[35] G. Mountrakis, J. H. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, 2011.
[36] S. K. Mylonas, D. G. Stavrakoudis, and J. B. Theocharis, "A GA-based sequential fuzzy segmentation approach for classification of remote sensing images," in *Proc. IEEE Int. Conf. Fuzzy Syst.* (FUZZ-IEEE), 2012, pp. 1–8.
[37] J. Nichol and M. S. Wong, "Remote sensing of urban vegetation life form by spectral mixture analysis of high-resolution IKONOS satellite images," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 985–1000, 2007.
[38] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 4, pp. 981–996, Jul. 1995.
[39] M. Pedergnana, P. Reddy Marpu, M. Dalla-Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Top. Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.
[40] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps Valls, J. Chanussot, M. Fauvel, P. Gamba, and A. Gualtieri, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, 2009.
[41] M. Salah, J. C. Trinder, M. Hamed, and A. Elsagheer, "Integrating multiple classifiers with fuzzy majority voting for improved land cover classification," in *Proc. Int. Arch. of Photogramm. Remote Sens.*, Part 3A, vol. 38, N. Paparoditis, M. Pierrot-Deseilligny, C. Mallet, and O. Tournaire, Eds., International Society for Photogrammetry and Remote Sensing ISPRS, 2010.
[42] B. Schölkopf, "The kernel trick for distances," *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 301–307, 2001.
[43] S. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.
[44] J. Serra, "Introduction to mathematical morphology," *Comput. Vis. Graph. Image Process.*, vol. 35, no. 3, pp. 283–305, 1986.
[45] G. Shao and J. Wu, "On the accuracy of landscape pattern analysis using remote sensing data," *Landsc. Ecol.*, vol. 23, no. 5, pp. 505–511, 2008.
[46] C. Smith Ramirez, "The Chilean coastal range: A vanishing center of biodiversity and endemism in South American temperate rainforests," *Biodivers. Conservat.*, vol. 13, no. 2, pp. 373–393, 2004.
[47] P. C. Smits, "Multiple classifier systems for supervised remote sensing image classification based on dynamic classifier selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 801–813, Apr. 2002.
[48] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation, and classification of hyperspectral data using watershed," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.* (IGARSS 2008), 2008, vol. 3, pp. III-652–III-655.
[49] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
[50] P. M. Treitz and P. J. Howarth, "Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems," *Progr. Phys. Geogr.*, vol. 23, no. 3, pp. 359–390, 1999.
[51] J. Trinder, M. Salah, A. Shaker, M. Hamed, and A. Elsagheer, "Combining statistical, and neural classifiers using Dempster-Shafer theory of evidence for improved building detection," in *Proc. 15th Australas. Remote Sens. Photogramm. Conf.*, Alice Springs, Australia, 2010, pp. 13–16.
[52] S. Velasco Forero and V. Manian, "Improving hyperspectral image classification based on graphs using spatial preprocessing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.,* (IGARSS 2008), vol. 3, 2008, pp. III-656–III-659.
[53] S. Velasco Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.
[54] J. D. Wegner, R. Hansch, A. Thiele, and U. Soergel, "Building detection from one orthophoto and high-resolution InSAR data using conditional random fields," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 83–91, Mar. 2011.
[55] D. C. Whitley, M. G. Ford, and D. J. Livingstone, "Unsupervised forward selection: A method for eliminating redundant variables," *J. Chem. Inf. Comput. Sci.*, vol. 40, no. 5, pp. 1160–1168, 2000.
[56] D. H. Wolpert and W. G. MacReady, "No free lunch theorems for optimization," *IEEE Trans. Evol. Comput.*, vol. 1, no. 1, pp. 67–82, Apr. 1997.
[57] J. Zhu and T. Hastie, "Kernel logistic regression and the import vector machine," *Adv. Neural Inf. Process. Syst.*, vol. 14, pp. 1081–1088, 2001.

Andreas Christian Braun was born in 1982. He received the Diploma in geoecology from the Technical University of Karlsruhe, Karlsruhe, Germany, in 2009, and the Ph.D. degree from the Karlsruhe Institute of Technology KIT, Karlsruhe, Germany. His diploma thesis is on vegetation classification with advanced machine learning algorithms. His Ph.D. thesis is on methods for the extraction of information on landuse change processes regarding vegetation using multispectral data. These landuse maps are combined with in situ assessments on biodiversity to regionalize human-induced alteration processes which lead to a biodiversity decline. In 2011, he was granted a funding for a feasibility study on environmental monitoring of landscape transformation processes in South America.

His main research interest are monitoring of environmentally relevant processes (such as deforestation or urban sprawl) by integration of geoecological field data and remotely sensed data. A focus of his geoecological research are human-induced alterations to ecosystems with special emphasis on biodiversity and habitat fragmentation.

Carolina Rojas received the Ph.D. degree in geographic information systems (GIS) and remote sensing from the University of Alcalá, Madrid, Spain.

She is a Chilean Geographer of Pontifical Catholic University of Valparaíso, Valparaíso Region, Chile. She is an Associate Professor at the Geography Department, University of Concepción, Biobío Region, Chile. Her work includes teaching and research in GIS, land uses and transport, and urban geography. Since 2011, she has been the Coordinator of the Masters Program in geographical analysis and from 2013, she is the main editorial board of Revista Geográfica de Sur. Her research is focused in biodiversity effects of urban growth and sprawl, urban sustainable development and transport, and urban mobility in Chilean cities. Currently, she is an Associate Researcher in the Center of Urban Sustainable development (CEDEUS), Santiago, Chile.

Cristian Echeverria received the degree of forest engineer and the M.Sc. degree at the Universidad Austral de Chile, Los Ríos Region, Chile. Then, he received the M.Phil. and Ph.D. degrees at the University of Cambridge, Cambridge, U.K.

Currently, he is the Head of Department of Forest Managment and Environment, Universidad de Concepcón, Biobío Region, Chile, and teaches courses on biodiversity conservation, landscape ecology, landscape modelling, and environmental management. His main contributions have been on the assessment of the pattern and impacts of forest loss and fragmentation in temperate and dryland landscapes. He leads the landscape ecology lab and is Chair Deputy for the International Union of Forest Research Organizations (IUFRO)-Landscape Ecology Working party.

Franz Rottensteiner received the Dipl.Ing. degree in surveying and the Ph.D. degree and venia legendi in photogrammetry, all from Vienna University of Technology, Vienna, Austria.

He worked as a Postdoctoral Researcher at the Vienna University of Technology, Vienna, Austria, University of New South Wales, Kensington, Australia, and University of Melbourne, Parkville, Australia. Currently, he is with the Institute of Photogrammetry and Geoinformation, Leibniz University of Hanover, Hanover, Germany, where he is the leader of the Photogrammetric Image Analysis Group. His research interests are digital photogrammetry, automated extraction of topographic objects, processing of lidar data, and sensor orientation.

Hans-Peter Bähr received the Dipl. Ing. and Dr. Ing. habil. degrees in geodesy from the University of Hannover, Hannover, Germany.

From 1981 to 1983, he worked at the Institution Building in Universidade Federal do Paraná (UFPR), Curitiba and in Universidade Federal de Pernambuco (UFPE), Recife, Brazil. From 1983 to 2008, he was the Head of the Institute of Photogrammetry and Remote Sensing (IPF), Karlsruhe, Germany. From 1994 to 2002, he was the President of German Geodetic Commission. He received the honorary Dr.h.c. in 2001 (Bucuresti) and worked as a Commander in Ordem da Sociedade Brasileira de Cartografia in 2005. His main research fields are in geometry of RS imagery, visual perception versus image understanding. He designed the graduate programs and lectured in UFPR, Curitiba, Brazil, and Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brazil, and University of Concepción (UdeC), Concepción, Chile. His research in these institutions included land-use analysis and cadastro multifinalitário.

Joachim Niemeyer studied geodesy and geoinformatics at Leibniz Universität Hannover, Hanover, Germany, and received the Dipl.Ing. degree, in 2009, where he is working toward the Ph.D. degree about the contextual classification of LiDAR data in urban areas.

Since 2010, he has been a Researcher at the Institute of Photogrammetry and GeoInformation (IPI), Leibniz Universität Hannover, Hannover, Germany. His research interests are laser scanning, laser bathymetry, and automated object detection.

Mauricio Aguayo Arias was born in 1972. He received the Ph.D. degree in environmental science from the Universidad de Concepcion, Concepcion, Chile.

He is a Forestry Engineer and Assistant Professor of the Environmental Science Faculty, University of Concepción, Biobío Region, Chile. His main research interests are ecological and environmental effects for landuse changes and spatially explicit modeling of the landuse change at the landscape scale. Between 2008 and 2011, he was a Young Scientist at the ANILLO scientific research project, social and environmental impacts of global climate change in the Biobío Region, challenges for the XXI century sustainability. Since 2007, he has led a number of scientific research projects funded national fund for scientific and technological development (FONDECYT) of Chile. He maintains a strong scientific collaboration with the Helmholtz Center for Environmental Research (UFZ), Leipzig-Halle, Germany. He has been in research stay in the Department of Ecological Modelling of the Helmholtz Center for Environmental Research (UFZ), developing issues of modeling patterns spatial. He has also led several projects related to the development and implementation of management plans for the conservation priority areas for both public and private institutions by establishing strong ties with the nonscientific community.

Mr. Arias has generated a range of scientific publications and training courses on environmental effects of the landuse change, spatially explicit modeling of the landuse patterns and scenarios future, and, in the last time, ecosystem services mapping.

Sergey Kosov was born in Mirnyy, Kazakhstan, in 1982. He received the Diploma degree in applied mathematics from the Kirgiz-Russian Slavic University, Bishkek, Kirgyzstan, in 2004, and the M.Sc. degree in computer science from the Saarland University, Saarbruecken, Germany, in 2008.

His current research interests include motion estimation with optical flow, three-dimensional stereo reconstruction, multigrid methods, classification with conditional random fields and software, as well as movie industry.

Stefan Hinz was the Head of a Helmholtz-Young-Investigator-Group and served as Scientific Secretary and Co-chair of several ISPRS working groups.

He is currently the Principle Investigator and Project Leader of numerous scientific projects dealing with model-based understanding of remote-sensing data. In addition, he is Speaker of KIT's Graduate School of Climate and Environment (GRACE), Karlsruhe, Germany. He has carried out research activities in remote sensing and computer vision, since 1998. From 1998 to 2003, his work focused on the development of theory and methods for the automatic extraction of linear infrastructural objects—mainly road networks in urban and rural environments—from single and multiview aerial or spaceborne optical images. As team leader at the remote sensing technology department of Technische Universität (TU) München, Munich, Germany (2004 onward), his research interests concentrated on automated image understanding using remote-sensing images. Special focus was on model-based methods for detection and characterization of dynamic processes with particular emphasis on traffic flow in urban environments. In this context, his group developed image analysis models for different sensors such as optical, infrared, and synthetic apeture radar (SAR) data of different scales. Since 2006, research activities extended to methods for automatic analysis of high-resolution optical and interferometric SAR images—in particular TerraSAR-X—with focus on Persistent Scatterer SAR Interferometry and high-resolution SAR simulation of urban environments.

Prof. Hinz has published more than 140 papers since 1998, of which more than 50 are peer-reviewed articles.

Uwe Weidner was born in 1963. He received the Dipl.Ing. degree in geodesy and the Dr.Ing. degree for his Ph.D. work on automatic building extraction from digital surface models from the Rheinische Friedrich-Wilhelms Universität, Bonn, Germany, in 1992 and 1997, respectively.

From 1997 to 2000, he worked for a company in the field of aerial photogrammetry and in 2000–2002, for a company in the field of industrial image processing and robotics. Since February 2003, he is with the Institute of Photogrammetry and Remote Sensing, KIT, Karlsruhe, Germany. His research interests include various of aspects automatic object extraction using in multi/hyperspectral remote-sensing data and evaluation of methods in remote sensing.