Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Fusion

Björn Waske, *Associate Member, IEEE*, and Sebastian van der Linden

***Abstract*—A strategy for the joint classification of multiple segmentation levels from multisensor imagery is introduced by using synthetic aperture radar and optical data. At first, the two data sets are separately segmented, creating independent aggregation levels at different scales. Each individual level from the two sensors is then preclassified by a support vector machine (SVM). The original outputs of each SVM, i.e., images showing the distances of the pixels to the hyperplane fitted by the SVM, are used in a decision fusion to determine the final classes. The fusion strategy is based on the application of an additional classifier, which is applied on the preclassification results. Both a second SVM and random forests (RF) were tested for the decision fusion. The results are compared with SVM and RF applied to the full data set without preclassification. Both the integration of multilevel information and the use of multisensor imagery increase the overall accuracy. It is shown that the classification of multilevel–multisource data sets with SVM and RF is feasible and does not require a definition of ideal aggregation levels. The proposed decision fusion approach that applies RF to the preclassification outperforms all other approaches.**

***Index Terms*—Data fusion, multilevel classification, multisensor data, random forests (RF), support vector machines (SVM), synthetic aperture radar (SAR).**

I. Introduction

THE RECENT development in land-cover classification of remote sensing images was to a great extent driven by three factors: 1) the increased availability of data from different, often complementary, sensors and sources; 2) the shift from statistical approaches to more powerful and flexible machine-learning algorithms for data classification; and 3) the introduction of operational image segmentation algorithms that allow the processing of data sets at multiple scales. These three aspects are interdependent and closely related to the parallel increase in computing power and data size. Richards [1] mentions the trend of enhanced numbers of bands and spectral resolution for optical sensors as well as the availability of multidimensional synthetic aperture radar (SAR) data at different wavelengths, polarizations, and incident angles as driving factors for the progress in digital image analysis methods. Due to limitations of earlier image data, many common methods use simple data models and techniques directly taken from the field of signal processing, e.g., maximum-likelihood estimation [1]. The rapidly increasing processing power impacts the pattern recognition algorithm development because it enables a faster processing of huge data sets. In addition, more stringent performance requirements like speed, accuracy, and cost demand more sophisticated methods [2]. As a result, the user can choose between several widely accepted algorithms and between a multitude of diverse remote sensing data. However, the greatest contemporary challenge might be the development of effective approaches for classifying multisensor imagery to bring together the inherent information contents of complementary data sets [1]. In general, the fusion of different data sources was discussed in several studies, for example, the fusion of multispectral imagery and topographical information for landslide monitoring [3], a wavelet-based technique for the fusion of multispectral and panchromatic imagery [4], or for reconstructing a digital terrain model under a woodland canopy [5]. In [6], a fuzzy fusion approach was used to combine multitemporal SAR data.

Multisensor data consisting of SAR and multispectral imagery positively impact the accuracies in various applications [7]–[12]. Although, in some studies, conventional methods like the well-known maximum-likelihood classifier are applied [7]–[9], such widely used statistical classifiers are often not optimal for these data sets. In most cases, the class distributions cannot be modeled by adequate multivariate statistical models [13]. Other multisource experiments are based on more sophisticated classification strategies and nonparametric algorithms [10]–[12]. In [10], a neural network and a statistical classifier are combined for classifying multisensor data. The SAR and multispectral data undergo a separate classification. Afterwards, the individual outputs are combined by decision fusion, which can be defined as a strategy for combining information from different data sources. In another study, Fauvel *et al.* [14] use decision fusion to improve the classification accuracy of high-resolution imagery from an urban area. Waske and Benediktsson [12] use a decision fusion strategy that is based on support vector machines (SVM) for the classification of SAR and multispectral imagery. SVM are a machine-learning development that has been successfully introduced in the remote sensing context [14]–[17] and still exhibits further improvement and modification [18]–[20]. The algorithm discriminates two classes by fitting an optimal separating hyperplane (OSH) only to those training samples that are closest to the class boundaries [21]. Unlike studies where a single SVM is applied

Manuscript received September 26, 2007. This work was supported in part by the German Aerospace Center (DLR) and in part by the Federal Ministry of Economics and Technology (BMWi) under the project Enviland (FKZ 50EE0404) as well as an ESA Cat-1 proposal (C1P 3115).

B. Waske is with the Department of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, Iceland (e-mail: bwaske@uni-bonn.de).

S. van der Linden is with the Geomatics Laboratory, Humboldt-Universität zu Berlin, 10099 Berlin, Germany (e-mail: sebastian.linden@geo.hu-berlin.de).

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/TGRS.2008.916089

to multisource data [22], Waske and Benediktsson [12] first classify the two data sets separately with two individual SVM. The final class memberships are then determined by applying an additional SVM to the original outputs of the first two SVM (i.e., distances to the hyperplane). This decision fusion outperforms the fusion with conventional voting schemes as well as a single SVM applied to the full data set or other parametric and nonparametric methods (i.e., maximum likelihood, decision tree (DT), and boosted DT).

Other studies on multisource and high-dimensional data use classifier ensembles [23]–[26]. The general strategy of such classifier ensembles, like Breimans' DT ensemble random forests (RF) [27], is based on training the same classifier on resampled input data. The resulting set of independent classifiers is then combined to create the final map. Briem *et al.* [23] show that classifier ensembles always outperform single classifiers in terms of accuracy on multisource data containing multispectral imagery, SAR data, and topographical information. In [25], a classifier ensemble was successfully used to classify multitemporal SAR data. The overall accuracy was significantly increased compared to the results achieved by a single DT. The concept of RF was successfully applied to hyperspectral imagery using a limited training sample set [24]. Gislason *et al.* [26] combine Landsat MSS data with topographical information and achieve the best results using RF.

Another development in the classification of remote sensing imagery is that of segment-based approaches [28]–[31], where adjacent pixels with similar properties are aggregated into image segments. After the segmentation process, information on the segments' mean spectral or backscatter value, their texture and shape, and neighborhood relationships can be derived. These can be used during subsequent classification. Shackelford and Davis [29] successfully combine information from pixel and segment levels to differentiate classes that are spectrally similar at pixel level. However, the definition of adequate aggregation levels is critical. Song *et al.* [30] discuss how inaccurate segmentation decreases classification accuracy. Bruzzone and Carlin [31] discuss the simultaneous use of multiple segmentation results at different scales as a feature-extraction module that adaptively models the spatial context of each pixel. They show that different levels contribute different types of information. Multilevel approaches with adequate classifier algorithms that combine relevant information from different aggregation levels appear to make the processing of information at different segmentation scales more effective. Segment-based approaches seem particularly interesting for agricultural areas that are dominated by typical spatial patterns of planted crops. Plot internal variations in spectral reflectance or backscatter intensity, which might be caused by differences in soil moisture, plant infections, etc., will be leveled out during image segmentation. Regarding SAR imagery and the inherent speckle in the data, the averaging of backscatter values reduces a significant amount of noise [28].

In this paper, the SVM-based decision fusion strategy in [12] is extended by a multilevel component. The concept is applied to a diverse data set from an agricultural area, consisting of multitemporal SAR data and a multispectral image. The two image sources are separately preclassified with SVM. In addition to the fusion with a second SVM, the process is also performed using RF. This strategy is assumed to make good use of both the ability of SVM to model complex class distributions in a higher dimensional feature space and the strength of RF to utilize various combinations of input features for optimized decision making. The results are compared to those achieved by a single SVM and a single RF on the full multisource–multilevel data set and on subsets of the data.

This paper is organized as follows. In Section II, an introduction to the conceptual framework of classifier ensembles and SVM is given. The data sets and preprocessing are described in Section III. The actual methods for the application of the proposed decision fusion strategy are explained in Section IV. Results are presented in Section V. Section VI discusses the results and concludes this paper.

II. Classification Algorithms

A. Classifier Ensembles—Random Forests

It has been shown in theory and with experimental results that the classification error can be decreased by combining different independent classifiers [32], [33]. The strategy is based on the assumption that independent classifiers produce individual errors, which are not produced by the majority of the other classifiers. In contrast to multiple classifiers that combine different methods [10], classifier ensembles like RF are based on independent variations of the same algorithm, i.e., the base classifier. DTs are used in most classifier ensemble approaches because they seem particularly interesting with their simple handling and fast training times [34]. Various strategies for generating independent outputs of the same classifier have been introduced, like the resampling of the training data (i.e., bagging or boosting) or feature space (e.g., random feature selection). In [35], an overview of different approaches is given.

Breiman's RF [27] build on DT classifiers $\{\mathrm{DT}(x, \theta_k), k = 1, \ldots\}$, where θ_k's are independent identically distributed random vectors and x is an input pattern. Each tree within the ensemble is trained on a bootstrapped sample (i.e., subset) of the original training samples. Moreover, the split rule at each split is determined by using only a randomly selected feature subset of the input data. A simple majority vote is used to create the final classification result.

The size of the feature subset is generally user defined, and the parameter is often set to the square root of the number of input features [26]. By decreasing the number of input features at each node, the computational complexity is simplified. Moreover, the correlation between the trees and, consequently, the corresponding outputs is decreased. This enables RF to handle high-dimensional data sets. Because the approach only employs subsets of the input data, the algorithm is considerably lighter than conventional DT-based bagging and boosting concepts [26].

In general, a DT classifier is based on a multistage or hierarchical concept. During the training process, the training data are sequentially partitioned into smaller increasingly homogenous subclasses using a set of rules defined at each split within the tree. Different methods for the determination of such split

rules have been introduced [36], which are usually based on the measurement of impurity (or rather purity) of the data within potential nodes. The RF algorithm uses the *Gini index*, a measure describing the impurity for a given node to find the largest homogeneous category within the training data and discriminate it from the remaining data [36]. The impurity of all subclasses is summed, and the split that causes the maximum reduction in impurity is chosen.

The gini index is described as

$$\text{Gini}(t) = \sum_{i=1}^{L} p_{\omega i}(1 - p_{\omega i}) \tag{1}$$

where $p_{\omega i}$, which is the probability or the relative frequency of class ω_i at node t, is defined as

$$p_{\omega i} = \frac{j_{\omega i}}{j} \tag{2}$$

with $j_{\omega i}$ as the number of samples belonging to class ω_i and j as the total number of samples within the training set.

B. Support Vector Machines

SVM differentiate two classes by fitting an OSH to the training samples of two classes in a multidimensional feature space. The optimization problem is based on structural risk minimization and aims to maximize the margins between the hyperplane and the closest training samples—the so-called support vectors [21]. Thus, SVM only use samples that are close to the class boundary. They perform efficiently with small training sets even when classifying high-dimensional imagery [37], [38]. For linearly not separable cases, the input data are mapped into a high-dimensional space, wherein the newly distributed samples enable the fitting of a linear hyperplane.

In recent studies, alternative methods to solve the optimization problem were presented, and the SVM approach was applied to semisupervised and ill-posed classification problems [18], [19]. Chi and Bruzzone [19] have introduced SVM implementations that are defined in the primal formulation problem of the optimization problem. Bazi and Melgani [20] combine genetic algorithms with SVM for an optimized classification system for hyperspectral data.

A detailed introduction on the general concept of SVM is given by Burges [39] and Schölkopf and Smola [40]. An overview in the context of remote sensing is given by Huang *et al.* [15] and Foody and Mathur [16].

For a binary classification problem in a d-dimensional feature space $\Re^d$, $x_i \in \Re^d$, $i = 1, 2, \ldots, L$, is a training set of L samples with the corresponding class labels $y_i \in \{1, -1\}$. The hyperplane $f(x)$ is described by a normal vector $w \in \Re^d$ and the bias $b \in \Re$, where $|b|/\|w\|$ is the distance between the hyperplane and the origin, with $\|w\|$ as the Euclidean norm from w

$$f(x) = w \cdot x + b. \tag{3}$$

The support vectors are located on two hyperplanes $w \cdot x + b = \pm 1$, which are parallel to the OSH.

The margin maximization leads to the following optimization problem:

$$\min \left[\frac{w^2}{2} + C \sum_{i=1}^{L} \xi_i \right] \tag{4}$$

with ξ_i as the slack variables and C as the regularization parameter that are introduced to handle misclassified samples in nonseparable cases. The constant C is added as a penalty for cases that lay on the wrong side of the hyperplane. Effectively, it controls the shape of the solution and thus affects the generalization capability of the SVM, e.g., a large value of C might cause an overfitting to the training data.

By using the so-called kernel methods, the aforementioned linear SVM approach is extended for nonlinear separable cases. During the nonlinear mapping, the data are transferred by a kernel function Φ into a higher dimensional feature space, and an OSH, which appears nonlinear in the original feature space, can be fit to a more complex class distribution. The input sample x can be described by $\Phi(x)$ in the new high dimensional space

$$(\Phi(x_i)\Phi(x_j)) = k(x_i, x_j). \tag{5}$$

Consequently, the final hyperplane decision function can be defined as

$$f(x) = \left(\sum_{i=1}^{L} \alpha_i y_i k(x_i, x_j) + b \right) \tag{6}$$

where α_i's are Lagrange multipliers.

The kernel trick enables one to work within the newly transformed feature space, without explicitly knowing Φ but only the kernel function. Widely used kernel functions are the polynomial and the Gaussian radial basis function kernel [21], [40]

Polynomial Kernel

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^a \tag{7}$$

Gaussian kernel

$$k(x_i, x_j) = \exp\left[-\gamma \|x_i - x_j\|^2\right] \tag{8}$$

where a denotes the order of the polynomial function and γ is the spread of the Gaussian kernel.

The training process involves the estimation of the kernel parameter γ and the regularization parameter C. In the literature, several approaches for an automatic model selection have been introduced, which employ a leave-one-out cross-validation procedure [41], [42].

Contrary to DTs, which directly provide a class label as output, the primary output images of SVM provided by (6) contain the distance of each pixel to the hyperplane of the binary classification problem, from now on referred to as rule images. These rule images are usually used to determine the final class membership.

SVM were originally developed as binary classifiers. Two main strategies exist for the solution of multiclass problems: An n-class problem is divided into several binary problems. For the one-against-one (OAO) strategy, a set of $n(n-1)/2$ individual SVM is trained, one for each possible pair of classes. By using

TABLE I
MULTITEMPORAL SAR DATA SET CONSISTING OF ENVISAT ASAR AND ERS-2 IMAGERY

SENSOR	IMAGE CHARACTERISTICS SAR DATA			
	Date	*Track / Swath*	*Polarization*	*Orbit*
ASAR	12-Apr-05	6208	HH / HV	Asc
ERS-2	21-Apr-05	337	VV	Des
ERS-2	26-May-05	337	VV	Des
ERS-2	30-Jun-05	337	VV	Des
ASAR	13-Jul-05	3029	HH / HV	Asc
ASAR	22-Jul-05	7158	HH / HV	Asc
ERS-2	4-Aug-05	337	VV	Des
ASAR	14-Aug-05	2487	HH / HV	Asc
ASAR	18-Sep-05	2487	HH / HV	Asc

the $n(n-1)/2$ rule images, a simple majority vote is applied to compute the final class membership. In case of the one-against-all (OAA) strategy, a set of n binary classifiers is trained to separate each class from the remaining ones, and the maximum decision value within the rule images is used to define the class label [21], [40].

In other approaches, the SVM classifier is directly described as a multiclass problem [43], [44]. However, a differentiation of more classes results in a more complex optimization problem [44], which might be instable compared to conventional binary optimization problems. Melgani and Bruzzone [17] have also introduced an alternative hierarchical tree-based strategy.

III. STUDY SITE AND DATA

The multisource data for the experiments include multitemporal SAR data (Table I) and one Landsat-5 Thematic Mapper (TM) image from 2005. Orthorectification and radiometric correction were applied to the multispectral image from May 26, 2005. The SAR imagery was calibrated to backscatter intensity following common procedures. The SAR-specific speckle noise was reduced by applying an enhanced frost filter. Finally, the SAR images at 12.5 m ground resolution were resampled to the pixel size of the Landsat scene (30 m) and orthorectified with a spatial accuracy of approximately one pixel using a digital elevation model, orbit parameters, and the corrected Landsat data as reference images.

The nearly flat test site is located near Bonn, Germany. The region is dominantly used for agriculture and characterized by typical spatial patterns caused by differences in the phenology of planted crops. The size of the field plots is around 5 ha. Cereals and sugar beets are the main crop types. An extensive ground truth campaign was conducted in summer 2005. The field data are used as a reference for the generation of training and validation data sets.

IV. METHODS

A. Image Segmentation

Several approaches for image segmentation have been introduced [45], [46]. In this paper, a frequently applied approach was used [29]–[31] that is based on a region-growing algorithm [46]. A detailed description of the concept is given by Shackelford and Davis [29], and general constraints are described by Bruzzone and Carlin [31]. When the segmentation procedure starts, each pixel is considered a separate segment, and subsequently, adjacent segments are merged into new larger segments. For the merge of two segments, the difference Δh between the heterogeneity of the new segment (e.g., spectral or backscatter heterogeneity) and the heterogeneity of the constituent segments is calculated as follows:

$$\Delta h = \sum_{i=1}^{b} W_i \left(\delta_1(h_{mi} - h_{1i}) + \delta_2(h_{mi} - h_{2i})\right) \qquad (9)$$

where h_{mi} is the heterogeneity of the potentially merged segment for the heterogeneity measure i (e.g., the variance in a specific band), and h_{1i} and h_{2i} are that of the two individual segments. The number of pixels belonging to the segments is denoted by δ_1 and δ_2. The weight W_i controls the impact of the specific heterogeneity. The merge is performed when this increase is below a user-defined heterogeneity criterion h_s. The segmentation process stops as soon as this condition cannot be met by any possible merge. Segment size is, hence, driven by h_s. In general, Δh depends on the spectral/backscatter variance within the user-defined bands and on the compactness or smoothness of segment shape. In this paper, only the spectral/backscatter variance was used, and all bands (i.e., Landsat bands and SAR acquisitions) were equally weighted.

Three independent image segmentations ($a1$, $a2$, and $a3$) with increasing values for h_s were performed on both the multispectral and the SAR data (Fig. 1). For both data sets, the average segment sizes are between approximately 10 and 55 pixels. Whereas segments at the smallest scale $a1$ contain only a few pixels and, normally, only fractions of features, outlines of the largest aggregation level $a3$ rather correspond to the largest agricultural plots. Further aggregation merges several natural features into single image segments. These levels do not provide additional useful information and, consequently, were not included.

B. Classification and Decision Fusion

Eight land-cover classes were considered for the experiment: arable crops, cereals, forest, grassland, orchard, rapeseed, root crops, and urban. One hundred fifty samples per class were selected from the ground mapping by equalized random sampling, guaranteeing that all eight classes are included in the sample set. These samples were used during both stages of the classification approach. Again, by using the equalized random sampling, an independent validation set with 500 samples per class was generated.

The SVM were trained individually on each aggregation level and the original pixel level of the SAR data and the multispectral image. The OAO strategy was followed to solve the multiclass problem using a set of binary classifiers with individually optimized parameters for γ and C. Although the OAO approach involves a larger number of binary SVM, the

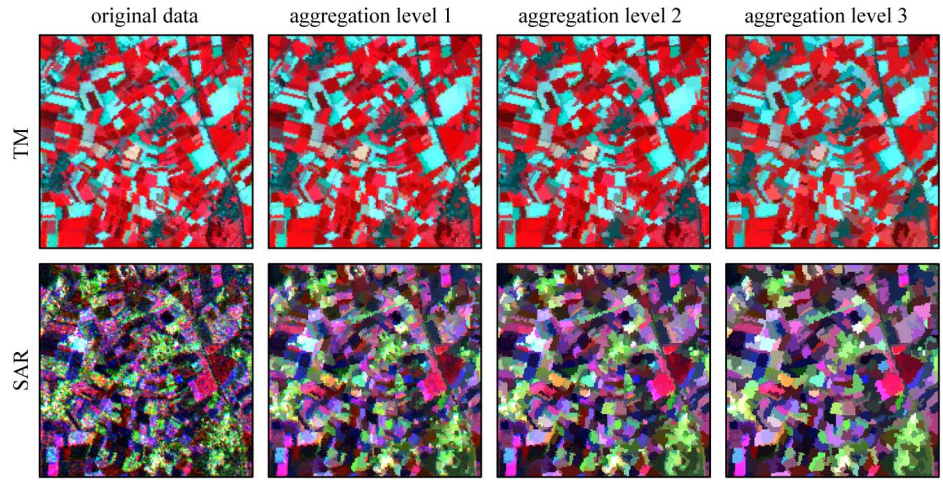

Fig. 1. Original data and corresponding aggregation levels of the multispectral TM (4/3/2) and multitemporal SAR (June 30/May 26/April 21) data.

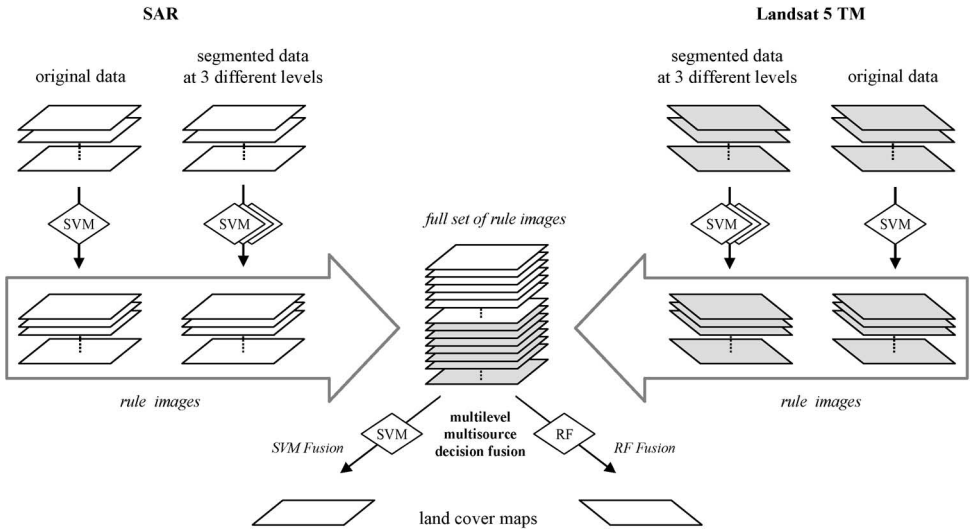

Fig. 2. Schematic diagram of the decision fusion strategy for multilevel imagery from different sensors.

classification problem as a whole is represented by simpler hyperplanes. This setup appears better suited to be used in a multiple classifier system than rule images from the OAA strategy that results in a smaller number of more complex hyperplanes. First tests proved this assumption.

In this paper, a Gaussian kernel was selected, which was used in several remote sensing studies. Contrary to a simple linear kernel, which is a special case of the Gaussian kernel, the Gaussian kernel can handle more complex nonlinear class distributions [47]. The polynomial kernel, on the other hand, requires more parameters than a Gaussian kernel, and the computational complexity of the model is increased. The training of the SVM with the Gaussian kernel and the generation of the rule images were performed using imageSVM [48]. The imageSVM is a freely available IDL/ENVI implementation, which uses the LIBSVM approach (http://www.csie.ntu.edu.tw/~cjlin/libsvm.) for the training of the SVM. The kernel parameters of C and γ are determined by a grid search using a tenfold cross validation. Possible combinations of C and γ are tested in a user-defined range, and the best combinations for γ and C were selected based on the tenfold cross validation.

TABLE II
OVERALL ACCURACY (IN PERCENT) USING SINGLE SVM AND SINGLE RF ON DIFFERENT DATA SETS AND SEGMENTATION LEVELS

AGGREGATION LEVEL	SAR		LANDSAT 5 TM	
	SVM	*RF*	*SVM*	*RF*
a0 (pixel)	63.3	65	69.8	68.8
a1	73.3	75.6	73.9	74.8
a2	75	78.1	75.5	77.2
a3	77.8	79.7	75.3	77.6
all levels	75.9	79.9	74.5	77.7

A decision fusion approach for the classification of multilevel data from two different sources is presented (Fig. 2). As a first step after preprocessing and image segmentation, a preclassification is performed. Individual SVM are trained on the original unsegmented data and each aggregation level of the two data sets, following the method described earlier. Each of these SVM classifiers provides 28 common rule images, given the eight land-cover classes.

TABLE III
CLASS-SPECIFIC ACCURACIES (IN PERCENT) AND CORRESPONDING STANDARD DEVIATION USING THE SINGLE SVM ON DIFFERENT DATA SETS AND SEGMENTATION LEVELS

LAND COVER CLASS	SAR				TM			
	PRODUCER ACCURACY		USER ACCURACY		PRODUCER ACCURACY		USER ACCURACY	
	a2	*a3*	*a2*	*a3*	*a2*	*a3*	*a2*	*a3*
Arable crops	72.4	72.8	71.7	83.1	71.2	81.6	79.1	70.2
Cereals	79.8	80.2	63.1	60.1	81.2	74.4	67.6	65.6
Forest	88.2	91.0	87.9	96.2	93.4	92.2	96.5	96.9
Grassland	65.8	65.4	77.4	70.8	55.8	62.0	71.7	70.3
Orchard	75.4	77.8	66.4	72.6	62.6	63.0	58.0	67.2
Rapeseed	70.8	75.6	85.7	88.1	73.8	70.6	82.4	77.6
Root crops	71.6	76.4	71.5	71.7	80.2	74.8	66.5	70.7
Urban	76.2	83.2	83.9	90.0	86.2	84.0	88.9	87.1
standard deviation	*6.8*	*7.5*	*9.2*	*12.2*	*12.3*	*10.4*	*12.8*	*11.0*

In conventional SVM classifications, each set of rule images would be used to determine the final result using a simple majority vote. For the proposed fusion strategy, all information from the rule images of all preclassifications is simply combined into one data set. In doing so, feature vectors are generated that contain the preliminary outputs of all aggregation levels of the two data sets. To determine the final multilevel–multisensor classification result, an additional SVM is applied to these feature vectors, which is analogous to that of Waske and Benediktsson [12] (from now on referred to as SVM fusion). Similar as done in the SVM fusion, RF were trained on the preliminary outputs of the SVM classifiers (from now on referred to as RF fusion). For the RF classification, a freely available Fortran code was used (http://www.stat.berkeley.edu/~breiman/RandomForests/).

For matters of comparison and to assess the quality of the multilevel and multisource fusion strategies, conventional SVM with OAO strategy and conventional RF were applied to various data sets. In contrast to the SVM fusion and RF fusion, these classifiers were directly trained on the original data and not on the corresponding rule images (from now on referred to as single SVM and single RF). The methods were applied to the following: 1) each individual aggregation level; 2) multilevel stacks from either source; 3) single aggregation-level multisource data set; and 4) a stack of all aggregation levels from both sources.

V. EXPERIMENTAL RESULTS

A. Multilevel Results

The different aggregation levels of the two data sets were separately classified by using both a single SVM and a single RF. Experimental results clearly show the positive effect of image segmentation (Table II). Irrespective of the classification method and the data type, the accuracy is significantly increased. By using the coarse aggregation level $a3$, the overall accuracies increase by up to 14% and 9% compared to the classification results achieved on the original SAR data and Landsat image, respectively. Even the use of the lowest level $a1$ improves the overall accuracy of the TM image by 6% and that of the SAR data by 10%. Image segmentation is particularly efficient for the SAR data. While accuracies of the original SAR data are below that of the multispectral imagery, the results achieved on levels $a1$ and $a2$ are comparable. For level $a3$, the accuracies of the SAR data are higher. The application of the classifiers to stacked data sets, including the pixel and all aggregation levels of either source, does not further improve the results. In Tables III and IV, the class-specific accuracies of two aggregation levels are presented for the single SVM and single RF. Accuracies vary strongly by land-cover type, the classifier algorithm, and the data source. The results show that the definition of an ideal segmentation level for all classes is not possible.

B. Multilevel–Multisource Results

For the final classification, the various aggregation levels of the two sensors were combined by the introduced concepts of SVM fusion and RF fusion. For comparison, a single SVM and a single RF were trained directly on a stacked multilevel/multisensor imagery, i.e., not the SVM rule images from individual levels/sensors.

Results clearly demonstrate the positive impact of a synergetic use of multisensor imagery at different scales. The highest accuracies achieved on a single-source imagery are 79.9% on the SAR imagery and 77.7% for the Landsat TM scene (Table II). The classification of the combined data set (i.e., multisource–multilevel) achieved the overall accuracies of 81.1% or higher (Table V). The performance is further increased to 84.9% (RF fusion) by the proposed fusion strategies. The advantage of using a diverse imagery can also be observed for the class-specific accuracies (Table VI). Most of the producer and user accuracies are increased by the SVM fusion and RF fusion. In addition, it can be assessed that the results are more balanced. Whereas the class-specific accuracies of the single-source classifiers vary strongly and show a relatively

TABLE IV
CLASS-SPECIFIC ACCURACIES (IN PERCENT) AND CORRESPONDING STANDARD DEVIATION USING THE SINGLE RF ON DIFFERENT DATA SETS AND SEGMENTATION LEVELS

LAND COVER CLASS	SAR				TM			
	PRODUCER ACCURACY		USER ACCURACY		PRODUCER ACCURACY		USER ACCURACY	
	a2	*a3*	*a2*	*a3*	*a2*	*a3*	*a2*	*a3*
Arable crops	77.4	81.0	75.7	77.1	77.0	83.6	71.6	70.4
Cereals	81.6	76.4	66.9	64.4	76.2	72.6	74.4	73.5
Forest	90.4	92.8	86.3	95.5	92.8	92.6	96.1	97.9
Grassland	70.8	69.6	79.0	75.0	61.0	70.4	72.8	67.3
Orchard	80.2	81.8	70.5	73.2	71.6	66.2	62.9	70.9
Rapeseed	77.4	80.6	82.2	83.6	75.4	76.0	83.2	82.3
Root crops	72.0	71.8	81.3	83.3	77.2	74.0	70.8	74.0
Urban	75.2	83.6	88.7	90.9	86.4	85.6	89.8	87.9
standard deviation	*6.2*	*7.3*	*7.5*	*10.0*	*9.5*	*8.8*	*11.0*	*10.5*

TABLE V
OVERALL ACCURACY (IN PERCENT), MULTILEVEL–MULTISOURCE CLASSIFICATION USING DIFFERENT FUSION STRATEGIES

CLASSIFIER / FUSION STRATEGY	OVERALL ACCURACY
Single SVM	81.1
SVM Fusion	83.5
Single RF	83.6
RF Fusion	84.9

TABLE VI
CLASS-SPECIFIC ACCURACIES (IN PERCENT) AND CORRESPONDING STANDARD DEVIATION USING SVM FUSION AND RF FUSION

LAND COVER CLASS	SVM FUSION		RF FUSION	
	Prod. acc.	*User acc.*	*Prod. acc.*	*User acc.*
Arable crops	79.6	84.8	78.2	95.1
Cereals	81.4	78.8	85.2	76.3
Forest	94.8	96.9	94.6	97.7
Grassland	76.6	76.4	82.2	76.7
Orchard	81.6	78.7	83.0	81.2
Rapeseed	84.8	84.8	89.0	79.6
Root crops	80.8	78.6	78.8	85.3
Urban	88.2	89.4	88.4	92.3
standard deviation	*5.7*	*6.9*	*5.6*	*8.5*

high standard deviation (Tables III and IV), the outcomes of the multisource approaches are less variable, resulting in a reduced standard deviation (Table VI).

To assess the value of a multilevel approach, the RF fusion concept was also applied to the multisensor data using only individual scales (Table VII). As before, results clearly demonstrate the impact of image segmentation: Whereas the multisensor data set results in a total accuracy of 77.4% at the pixel level,

TABLE VII
OVERALL ACCURACIES (IN PERCENT), FUSING MULTISENSOR IMAGERY AT DIFFERENT AGGREGATION LEVELS

AGGREGATION LEVEL	RF FUSION
a0 (pixel)	77.4
a1	81.1
a2	83.8
a3	83.5
all levels	84.9

the image segmentation improves the results by 3.7% or more. However, the results from the proposed multisensor–multilevel approach that fuses all aggregation levels are always better than the single-level multisensor data.

The visual assessment of the classification results (Fig. 3) underlines the positive effect of image segmentation and multisensor fusion. Concerning the multilevel aspect, the noise within the pixel-based results is clearly reduced, in particular, the speckle in the SAR imagery. Plots generally appear more homogeneous, and some typical errors, e.g., the confusion of forest and urban in the SAR data, are eliminated. In addition, the complementary character of the multitemporal SAR and the multispectral optical data is shown by the results from the RF fusion approach: Resulting maps are also more homogeneous, and errors for critical classes are decreased. Despite the visibly best performance of the RF fusion on the multilevel–multisource data, the disappearance of the highway in the east, which was recognized in the TM data, shows possible drawbacks of the multisensor approach.

VI. DISCUSSION AND CONCLUSION

In the study presented here, the problem of classifying the multisensor imagery at different segmentation levels is addressed. Both aspects, the combination of information from

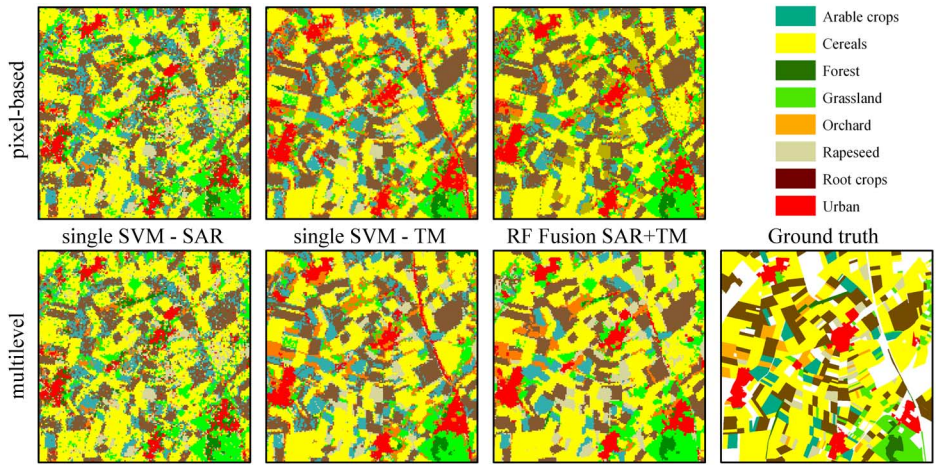

Fig. 3. Classification results of a single SVM and RF fusion applied to the original pixel-based imagery and the multilevel data set. The ground truth mapping is shown for comparison.

different sources and the work on multiple levels to integrate spatial context on different scales, prove useful.

Image segmentation alone already improves the results regardless of the aggregation level and sensor (Table II). In particular, the segmentation inherent speckle reduction significantly improves the accuracy for the SAR data and shows the quality of multitemporal data for the classification of agricultural areas. However, the high accuracy of the multilevel–multisource approach owes not only to the segmentation but also relates to the multisensor fusion: The classification of a stacked image containing SAR and TM data at the pixel level achieves an overall accuracy of 77.4%, fusing the preliminary outputs by RF. This is significantly higher than the results on single data sets at the pixel level. This positive impact of using multisensor data is also confirmed by the single-level multisensor results, which yield higher accuracies than the individual sensor sources at the same aggregation level. Nevertheless, the results achieved on the multilevel–multisource data are always better than those achieved on any other combinations of data. The multisensor results that employ each aggregation level outperform all single-scale results (Table VII).

The assessment of class-specific accuracies underlines the general increase in quality. Both classification algorithms, the SVM and RF, perform well on the combined data set. In particular, the single RF shall be mentioned as a simple yet accurate approach. Results are further improved by the decision fusion strategies suggested in this paper. When SVM or RF are trained on the rule images from the SVM preclassification of individual levels, accuracies outperform those from the corresponding single classifier. The RF fusion achieves the best overall accuracy. Results from the SVM fusion, on the other hand, show a very positive balance between producer and user accuracies of the classes.

To investigate the stability of the proposed concept, the approach was applied to a similar data set from year 2006. It was made up of three SPOT scenes and ten SAR acquisitions, including ERS-2 and ENVISAT ASAR data. The classification problem was extended by discriminating three additional land-cover classes. The results show the same characteristics as the ones using data from 2005 (Table VIII).

TABLE VIII
OVERALL ACCURACIES (IN PERCENT), FUSING MULTISENSOR IMAGERY FROM 2006

AGGREGATION LEVEL	RF FUSION
a0 (pixel)	75.6
a1	79.7
a2	80.5
a3	80.5
all levels	82.7

We assume the reason for the success of the presented fusion framework to be the combination of the two classifier methods. The sequential use of SVM and RF comprises the strengths of both algorithms: SVM can model complex class distributions, and RF consider various combinations of input features for an optimized decision making.

In this context, the preclassification by SVM can be regarded as a class-specific data transformation. By applying SVM with the OAO multiclass strategy, the image data are transformed into a new feature space that is made up of the distance values of the individual SVM rule images. The values of these rule images are better comparable than the original data sets, and each binary rule image is more appropriate to describe the interclass differences. This nonlinear class-specific data transformation simplifies the definition of the split rules during DT training and leads to the best results of the RF fusion. This is in accordance with the results in [49], where a DT is applied to the rule images of a spectral angle mapping. Results of the combined approach outperform those from a single spectral angle mapper and a single DT.

Regarding the experimental results, it seems more adequate to define the kernel functions for each data source and level separately, instead of using one specific kernel function on the full data set. Furthermore, the different nature of the image sources provides a diverse information, and the different aggregation levels contribute unequally to the classification of the various land-cover classes (Tables III and IV). The different

information levels/sources, hence, appear not equally reliable. Sometimes, one specific source (i.e., image type or specific aggregation level) is well suited to describe a class number one, whereas a second source is not ideal for this class. At the same time, it might be the other way around for class number two. Thus, it can be assumed that the individual errors are diverse and uncorrelated. This fact strengthens the performance of a multiple classifier system compared to a single classifier approach, and the RF strategy appears appropriate for decision making in this context.

The experimental results show that the useful combination of multilevel–multisensor data is feasible with contemporary machine-learning algorithms like SVM and RF. The sequential application of the two approaches in the suggested fusion approach further improves the results of the classification. The approach is assumed to be universally applicable for the work with multiple data sets but also useful for complex application with individual sources. By following this strategy, multilevel data can be processed without the definition of specific ideal segmentation levels. It is expected that the positive impact of multilevel approaches becomes even more dominant in the context of high-resolution imagery. For the work with more complex class distributions in single data sets, e.g., in hyperspectral data from urban areas, the ability of the SVM preclassification to function as a class-specific data transformation might have an even more positive impact. Hence, the proposed strategy seems particularly interesting with respect to upcoming EO missions.

ACKNOWLEDGMENT

The authors would like to thank V. Heinzel, M. Braun, and G. Menz from the Center for Remote Sensing of Land Surfaces, University of Bonn, for their support within the research project and field works. The authors would also like to thank the valuable comments of the anonymous reviewers and the guest editor.

REFERENCES

[1] J. A. Richards, "Analysis of remotely sensed data: The formative decades and the future," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 422–432, Mar. 2005.

[2] K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[3] Y.-L. Chang, L.-S. Liang, C.-C. Han, J.-P. Fang, W.-Y. Liang, and K.-S. Chen, "Multisource data fusion for landslide classification using generalized positive Boolean functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1697–1708, Jun. 2007.

[4] P. S. Pradhan, R. L. King, N. H. Younan, and D. W. Holcomb, "Estimation of the number of decomposition levels for a wavelet-based multiresolution multisensor image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 12, pp. 3674–3686, Dec. 2006.

[5] C. S. Rowland and H. Balzter, "Data fusion for reconstruction of a DTM, under a woodland canopy, from airborne L-band InSAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1154–1163, May 2007.

[6] J. Chanussot, G. Mauris, and P. Lambert, "Fuzzy fusion techniques for linear features detection in multitemporal SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1292–1305, May 1999.

[7] B. Brisco and R. J. Brown, "Multidate SAR/TM synergism for crop classification in western Canada," *Photogramm. Eng. Remote Sens.*, vol. 61, no. 8, pp. 1009–1014, 1995.

[8] G. Chust, D. Ducrot, and J. L. Pretus, "Land cover discrimination potential of radar multitemporal series and optical multispectral images in a Mediterranean cultural landscape," *Int. J. Remote Sens.*, vol. 25, no. 17, pp. 3513–3528, Sep. 2004.

[9] H. Huang, J. Legarsky, and M. Othman, "Land-cover classification using Radarsat and Landsat imagery for St. Louis, Missouri," *Photogramm. Eng. Remote Sens.*, vol. 73, no. 1, pp. 37–43, Jan. 2007.

[10] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

[11] D. B. Michelson, B. M. Liljeberg, and P. Pilesjö, "Comparison of algorithms for classifying Swedish landcover using Landsat TM and ERS-1 SAR data," *Remote Sens. Environ.*, vol. 71, no. 1, pp. 1–15, Jan. 2000.

[12] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[13] J. A. Benediktsson, P. H. Swain, and O. K. Ersoy, "Neural network approaches versus statistical methods in classification of multisource remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 4, pp. 540–552, Jul. 1990.

[14] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Decision fusion for the classification of urban remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2828–2838, Oct. 2006.

[15] C. Huang, L. S. Davis, and J. R. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.

[16] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.

[17] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

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

[19] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVM optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[20] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 3374–3385, Nov. 2006.

[21] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[22] X. Song, F. Fan, and M. Rao, "Automatic CRP mapping using nonparametric machine learning approaches," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 888–897, Apr. 2005.

[23] 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.

[24] J. Ham, Y. Chen, and M. M. Crawford, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[25] B. Waske, S. Schiefer, and M. Braun, "Random feature selection for decision tree classification of multi-temporal SAR data," in *Proc. IGARSS*, Denver, CO, 2006, pp. 168–171.

[26] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, Mar. 2006.

[27] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[28] B. Tso and R. M. Mather, "Crop discrimination using multi-temporal SAR imagery," *Int. J. Remote Sens.*, vol. 20, no. 12, pp. 2443–2460, Aug. 1999.

[29] A. K. Shackelford and C. H. Davis, "A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2354–2363, Oct. 2003.

[30] M. Song, D. L. Civco, and J. D. Hurd, "A competitive pixel-object approach for land cover classification," *Int. J. Remote Sens.*, vol. 26, no. 22, pp. 4982–4997, Nov. 2005.

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

[32] K. Turmer and J. Gosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recognit.*, vol. 29, no. 2, pp. 341–348, Feb. 1996.

[33] J. Kittler, "Combining classifiers: A theoretical framework," *Pattern Anal. Appl.*, vol. 1, no. 1, pp. 18–27, Mar. 1998.

[34] M. Pal and P. M. Mather, "An assessment of the effectiveness of decision tree methods for land cover classification," *Remote Sens. Environ.*, vol. 86, no. 4, pp. 554–565, Aug. 2003.

[35] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 173–180, Jan. 2007.
[36] M. Zambon, R. Lawrence, A. Bunn, and S. Powell, "Effect of alternative splitting rules on image processing using classification tree analysis," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 1, pp. 25–30, 2006.
[37] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Evaluation of kernels for multiclass classification of hyperspectral remote sensing data," in *Proc. IEEE ICASSP*, Toulouse, France, 2006, pp. II-813–II-816.
[38] M. Pal and P. M. Mather, "Some issues in the classification of DAIS hyperspectral data," *Int. J. Remote Sens.*, vol. 27, no. 14, pp. 2895–2916, Jul. 2006.
[39] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, Jun. 1998.
[40] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.
[41] O. Chapelle, V. Vapnik, O. Bousquet, and S. Mukherjee, "Choosing multiple parameters for support vector machines," *Mach. Learn.*, vol. 46, no. 1–3, pp. 131–159, Jan. 2001.
[42] K.-M. Chung, W.-C. Kao, C.−L. Sun, L.-L. Wang, and C.-J. Lin, "Radius margin bounds for support vector machines with the RBF kernel," *Neural Comput.*, vol. 15, no. 11, pp. 2643–2681, Nov. 2003.
[43] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
[44] D. J. Sebald and J. A. Bucklew, "Support vector machines and the multiple hypothesis test problem," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 2865–2872, Nov. 2001.
[45] C. Evans, R. Jones, I. Svalbe, and M. Berman, "Segmenting multispectral landsat TM images into field units," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 5, pp. 1054–1064, May 2002.
[46] M. Baatz and A. Schäpe, "Multiresolution segmentation—An optimization approach for high quality multi-scale segmentation," in *Angewandte Geogr. Informationsverarbeitung*, J. Strobel *et al.*, Ed. Karlsruhe, Germany: Wichmann, 2000, pp. 12–23.
[47] S. S. Keehrti and C.-J. Lin, "Asymptotic behaviors of support vector machines with Gaussian kernel," *Neural Comput.*, vol. 15, no. 7, pp. 1667–1689, Jul. 2003.
[48] A. Janz, S. Schiefer, B. Waske, and P. Hostert, "Image SVM—A user-oriented tool for advanced classification of hyperspectral data using support vector machines," in *Proc. 5th Workshop EARSeL Special Interest Group Imaging Spectroscopy*, Bruges, Belgium, 2007.
[49] S. van der Linden, B. Waske, and P. Hostert, "Towards an optimized use of the spectral angle space," in *Proc. 5th Workshop EARSeL Special Interest Group Imaging Spectroscopy*, Bruges, Belgium, 2007.

Björn Waske (S'06–A'08) received the M.Sc. degree in applied environmental sciences with a major in remote sensing from the University of Trier, Trier, Germany, in 2002 and the Ph.D. degree from the University of Bonn, Bonn, Germany, in 2007.

He is currently with the Department of Electrical and Computer Engineering, University of Iceland, Reykjavik, Iceland. From 2004–2007 he was with the Center of Remote Sensing of Land Surfaces (ZFL), University of Bonn. Until mid-2004, he was a Research Assistant with the Department of Geosciences, Munich University, Munich, Germany, where he worked on the use of remote sensing data for flood forecast modeling. In 2006, he visited the Department of Electrical and Computer Engineering, University of Iceland, for three months. His current research interests include advanced concepts for image classification and data fusion, focusing on multisensor applications.

Dr. Waske is a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Sebastian van der Linden received the M.Sc. degree in applied environmental sciences from the University of Trier, Trier, Germany, in 2002 and the Ph.D. degree from Humboldt-Universität zu Berlin, Berlin, Germany, in 2008.

He is currently a Postdoctoral Researcher with the Geomatics Laboratory, Humboldt-Universität zu Berlin. From 1999 to 2000, he studied at the University of Edinburgh, Edinburgh, U.K. In 2006, he spent three months as a Visiting Scholar with the Lamont-Doherty Earth Observatory, Columbia University, New York. He is a Reviewer for *Remote Sensing of Environment* and the *Canadian Journal of Remote Sensing*. His current research and teaching interests include qualitative and quantitative approaches for imaging spectroscopy data, focusing on the application of advanced processing methods.

Dr. van der Linden is a Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.