PFG 2013 / 3, 0149–0161
Stuttgart, June 2013

Article

Object Features for Pixel-based Classification of Urban Areas Comparing Different Machine Learning Algorithms

Nils Wolf, Bochum

Keywords: urban remote sensing, machine learning, object-based feature extraction, spatial-spectral classification

Summary: Image segmentation is a means to extract valuable spatial descriptors from urban remote sensing images. These descriptors (object features) can be concatenated at the pixel level to the spectral feature vector. Resulting spatial-spectral input spaces become more complex, but it is assumed that they offer in conjunction with appropriate classification techniques a better discrimination of classes. Comparing different supervised learning algorithms, this study empirically evaluates the value of adding object-based features into per-pixel classification. The considered algorithms are decision tree, decision tree ensembles (bagging and random forest), support vector machines (linear and rbf kernel), and k-nearest neighbour. The pixel level is suggested as the preferable domain for accessing object features in order to facilitate unbiased training, tuning and testing of algorithms within an implemented nested cross-validation scheme. Different case studies of urban remote sensing are considered to conduct the experiments, namely building detection with hyperspectral data (CASI) and aerial photography (Leica RC30), the mapping of pools, turf grass and non-turf vegetation in an urban tourist area using WorldView-2 panchromatic data, urban land cover classification using a hyperspectral benchmark dataset (ROSIS) and the classification of urban tree species (CASI). The results show that spatial features derived from segmentation levels have a great value for these applications. Concerning the algorithm performance, decision tree ensemble and support vector machine approaches yield in overall better results than decision tree and k-nearest neighbour.

Zusammenfassung: *Objektmerkmale für die pixelbasierte Klassifizierung urbaner Räume: ein Vergleich von Algorithmen des maschinellen Lernens.* Die Bildsegmentierung ermöglicht es, aus urbanen Fernerkundungsszenen wichtige räumliche Deskriptoren zu extrahieren. Diese Deskriptoren (Objektmerkmale) können auf Pixelebene mit dem spektralen Merkmalsvektor kombiniert werden. Resultierende räumlich-spektrale Merkmalsräume sind komplexer, aber es wird vermutet, dass sie in Verbindung mit angemessenen Klassifizierungstechniken eine bessere Trennung von Klassen ermöglichen. Im Vergleich verschiedener überwachter Lernalgorithmen untersucht diese Studie empirisch den Wert von objektbasierten Merkmalen für die pixelbasierte Klassifizierung. Die untersuchten Algorithmen sind decision tree, decision tree ensembles (bagging und random forest), support vector machines (linear und rbf kernel) und k-nearest neighbour. Die Pixelebene ist dabei als bevorzugte Domäne für das unverzerrte Trainieren, Tunen und Testen der Algorithmen innerhalb einer verschachtelten Kreuzvalidierung anzusehen. Für die Durchführung der Experimente werden verschiedene Fallstudien urbaner Fernerkundung berücksichtigt, namentlich Gebäudedetektion sowohl mit Hyperspektralscanner-Daten (CASI) als auch Luftbildern (Leica RC30), Kartierung von Swimmingpools, Rasenflächen und Baum-/Strauchvegetation in einer touristisch geprägten urbanen Region mit panchromatischen WorldView-2 Daten, Klassifizierung urbaner Landbedeckung mit einem hyperspektralen Benchmark-Datensatz (ROSIS) und die Klassifizierung urbaner Baumarten (CASI). Die Ergebnisse zeigen, dass die aus Segmentierungsebenen extrahierten räumlichen Merkmale für diese Anwendungsbeispiele einen bedeutenden Mehrwert haben. Hinsichtlich des Algorithmenvergleichs lieferten decision tree ensembles und support vector machines übergreifend deutlich genauere Ergebnisse als decision tree und k-nearest neighbour.

www.schweizerbart.de
DOI: 10.1127/1432-8364/2013/0166
1432-8364/13/0166 $ 3.25

1 Introduction

Classification is a common processing step to convert image data into tangible information and its practice in remote sensing is well established for a wide range of applications (Lu & Weng 2007). A particular problem in urban remote sensing is the spectral diversity (Herold et al. 2003) and the spatial complexity of this environment, which complicates decoding of the spectral signals. This is especially valid for studies mapping inner cities or targeting specific materials and objects. The complexity relates to the various artificial and natural surfaces and, moreover, to the presence of vertical structures and non-horizontal surfaces which cast shadows and effect the angular distribution of reflected light. This finally limits analysis approaches which solely rely on the pixelwise interpretation of the spectral values. In fact, these approaches treat an image as being an unordered list of spectral measurements and neglect the spatial structure of these measurements. This is the reason why there is wide consensus that spatial information considered in image analysis can be advantageous and thus, various methods have been developed and applied, for example those based on textural, morphological or object features (Lu & Weng 2007, Palmason et al. 2005, Tarabalka et al. 2010).

From about the year 2000 onwards, the object-based paradigm has gained increasing attention (Blaschke 2010), with a research focus on knowledge-driven approaches (Baatz & Schäpe 1999, Benz et al. 2004, Blaschke et al. 2008, Burnett & Blaschke 2003). As a consequence, supervised classification algorithms and comparison studies, which were well established in the pixel-based context, e.g. Huang et al. (2002) and Waske et al. (2009), have been rarely studied under the object-based paradigm. This is surprising because it can be advantageous or a complementary approach to complex classification problems where the knowledge representation and organization becomes difficult and loses its operational strengths of transparency and transferability, i.e. the independence of a scene.

Only recently, an increased interest in supervised object-based approaches employing state of the art machine learning algorithms can be observed. For example, Duro et al. (2012) compared the decision tree, support vector machine and random forest classifiers by mapping broad land cover categories in an agricultural landscape using SPOT HRG imagery and found out that random forests and support vector machines perform significantly better than decision trees. Novack et al. (2011) compared decision trees, regression trees, random forests and support vector machines for urban land cover classification using WorldView-2 as well as simulated QuickBird data and concluded that in overall random forests provided the best results and support vector machines the worst. However, one general problem of the application of supervised methods in the object domain is the fact that objects, unlike pixels, are of non-uniform size and distribution and, moreover, that they constitute a generalization of a raster image, i.e. they are even more than pixels affected by thematic uncertainties and ambiguities. Both has an impact on the representation of data samples, i.e. labelled cases used for training, tuning and testing which can misdirect the optimization procedures of algorithms and violate premises of statistically rigorous accuracy assessments (Stehman & Czaplewski 1998).

Having said this, the current paper presents five case studies in which object-based features are used in a pixel-based framework in order to apply and compare a number of non-parametric supervised machine learning algorithms. Object feature layers are extracted from multiple segmentation levels and then stacked as additional raster layers to the original input image. In this way, the spectral feature vector of each pixel is linked to a feature vector that carries spatial information. Using a nested cross-validation scheme, it is investigated whether the resulting *spatial-spectral* input spaces yield higher classification accuracies than the conventional *spectral* input spaces. The algorithms under consideration are decision trees, decision tree ensembles (bagging and random forest), support vector machines (linear and radial basis function kernels), and – to provide a low-cost benchmark – k-nearest neighbour.

2 Materials and Methods

2.1 *Case Studies: Data and Application*

Building detection with CASI hyperspectral data

Hyperspectral data of Bochum, Germany, were acquired on 5th July 2011 with the Compact Airborne Spectrometry Imager (CASI) mounted on an aircraft. The data are available with 1 m/pixel geometric resolution and 72 continuous bands in the range from 380 nm to 1050 nm. A 1341 ha test site (1381 × 9714 pixels) was chosen that covers different urban structures, including high-density block development, industrial and commercial areas, terraced houses, perimeter block development as well as parks and allotments. Using simple random sampling, 2000 pixels have been selected and labelled (*building* or *background*) by intersection with an official roof area cadastre map.

Building detection with colour aerial photography (Leica RC30)

The building detection was also conducted using aerial photography. The images were taken during a flight campaign on 2nd April 2009 with the analogue Leica RC30 camera (film, 400 nm – 1000 nm, colour filter) and later digitalized with a photogrammetric scanner to RGB layers in 10 cm/pixel resolution. To reduce the data volume, the pixel size was degraded to 40 cm. The 1341 ha test site (3,452 × 24,285 pixels) and the procedure to generate labelled sample data are identical to the CASI-based application (cp. above).

Mapping of pools, turf and tree/shrub vegetation with WorldView-2 (WV-2) panchromatic data

A WV-2 satellite scene of the vicinity of Sotogrande, Andalusia, Spain, recorded on 16th July 2010, was available. The data were used to map *swimming pools*, *turf grass* and *trees/shrubs*. The remaining area was considered as *background*. The chosen test site (502 ha subset, ≈ 20 million pixels) represents a high-quality, low-density tourist and residential area, dominated by large private properties with large garden area, most of them include a swimming pool. These urban-tourist zones most notably along the Spanish Mediterranean coast are particularly vulnerable to water shortages induced by climate change. Mapping water-intensive landscape features like pools and irrigated garden vegetation, especially turf, provides an important input for water consumption studies (Hof & Schmitt 2011). Labelled information was available from a visual interpretation of 2059 randomly distributed pixels. For the experiments, only the panchromatic band (0.5 m/pixel, 463 nm – 800 nm spectral range) was chosen in order to test the value of spatial information layers in a context where spectral information is limited.

Urban land cover classification with ROSIS hyperspectral data

Data with 115 spectrally continuous bands in the 430 nm – 860 nm spectrum were recorded (8th July 2002) by the Reflective Optics System Imaging Spectrometer (ROSIS) over Pavia, Italy. Two image subsets (together 715 × 1096 pixels, 102 ha) with 1.3 m/pixel spatial resolution and 102 bands (after removal of noisy bands) were available, covering the city centre. Nine classes, namely *water*, *trees*, *asphalt*, *bricks*, *bitumen*, *tiles*, *shadows* and *meadows* are considered for the experiment. The labelled data sample consists of 99,114 pixels which have been converted from sample polygons (nonprobability sampling). The ROSIS image is a well-known, publicly available benchmark dataset (Palmason et al. 2005, Tarabalka et al. 2010). It was chosen because it provides a good source for transparent algorithm benchmarking beyond the scope of a single study.

Tree species classification with CASI hyperspectral data

The imagery used here is another subset of the CASI data recorded during the flight campaign on 5th July 2011 over the city of Bochum. Labelled information was generated using an official tree cadaster of the local planning authorities. It provides the locations and the species information of trees in public space, e.g. on cemeteries, sports facilities, parks or along streets. In a defined 1471 ha subset (1470 × 6242 pixels), the cadaster was filtered to remove sparse species, maintaining only the eight most frequent ones which sum up to 400 trees. They are: *Acer platanoides* (28 cases)*, Acer pseudoplatanus* (171)*, Acer saccharinum* (43)*, Aesculus x carnea "Briotii"* (33)*, Betula pendula* (29)*, Crataegus x lavallei* (17)*, Platanus x acerifolia* (40) and *Tilia tomentosa* (39).

2.2 *Machine Learning*

This paper deals with supervised learning which is among several machine learning paradigms, like unsupervised, active and reinforcement learning, arguably the most relevant one in terms of impact on practical applications. The basic notion of supervised learning is the approximation of a target function f: $X \rightarrow Y$ based on observational or sample data $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, with x being a p-dimensional input vector and y an outcome. X and Y denote the complete set of possible input and output data and remaining unknowns, likewise f. Learning tasks can be categorized into regression leading to continuous output, and classification leading to discrete output. The empirical experiments and the following introductory descriptions address the classification case. The algorithms are implemented within the statistical programming environment *R* (R Core Team 2012) and its contributed packages. An overview of the algorithms with the related *R* packages, functions, parameters and key references (for implementation details) is given in Tab. 1.

Decision Tree (DT)

DTs recursively partition the input space by axis-parallel splits into a set of rectangular areas, aiming thereby at grouping data points with the same class. The implementation used in this work is based on classification and regression trees (CART) by Breiman (1984). Following a greedy problem solving heuristic, trees begin with a root node where the locally best univariate binary split is selected. The selection relies on an exhaustive search through all variables and possible thresholds regarding a defined measure which quantifies the reduction in class impurity obtained by a particular split. Then, the child nodes are considered themselves as the new roots and the process iterates until pure end nodes, the leaves, are reached. The impurity measure used in this study is the Gini index (Breiman 1984):

$$G = \sum_{i=1}^{c} P(\omega_i)(1 - P(\omega_i)) \tag{1}$$

Tab. 1: Overview of algorithms and their implementation and calibration (DT = Decision Tree, BAG = Bagging, RF = Random Forest, SVM = Support Vector Machines, rbf = radial basis function, KNN = k-Nearest Neighbour).

Algorithm	**R package/*function*** **key reference**	**Fixed parameter**	**Tuning parameter**
DT	rpart / *rpart*() (Therneau & Atkinson 1997)	n/a	$cp = \{2^{-10}, 2^{-9}, ..., 2^{-1}\}$
BAG	randomForest / *randomForest*() (Svetnik et al. 2003)	$ntree = 500$ $mtry = \mathrm{p}$	n/a
RF	randomForest / *randomForest*() (Svetnik et al. 2003)	$ntree = 500$ $mtry = \lceil \sqrt{p} \rceil$	n/a
SVM rbf	kernlab / *ksvm*() (Karatzoglou et al. 2004)	*kernel* = "rbfdot"	$C = \{5^{-2}, 5^{-1}, ..., 5^{7}\}$ $\sigma = \{5^{-7}, 5^{-6}, ..., 5^{-1}\}$
SVM linear	kernlab / *ksvm*() (Karatzoglou et al. 2004)	*kernel* = "vanilladot"	$C = \{5^{-2}, 5^{-1}, ..., 5^{7}\}$
KNN	class / *knn*() (Venables & Ripley 2002)	n/a	$k = \{2^{0}, 2^{1}, ..., 2^{4}\}$

with $P(\omega_i)$ being the relative frequency of the ith class out of c classes in the node. The impurity reduction G_{gain} of a particular split is derived by comparing the impurity of the current root node G_{root} with its child nodes G_{left} and G_{right}:

$$G_{gain} = G_{root} - (G_{left} + G_{right}). \tag{2}$$

The variable-threshold combination with the highest G_{gain} is chosen to split the data. Trees, as described so far, are fully grown in the sense that they branch out until pure end nodes are reached. Thereby, they overfit the training data (performing 100 % accurate on it) and miss generality when it comes to the prediction of unseen data. To restrict the complexity of trees, several *pruning* approaches exist. The one used in this study is based on the complexity cost *cp* (Therneau & Atkinson 1997), here defined by a tuning range (Tab. 1).

Bagging (BAG) and Random Forest (RF)

BAG is an ensemble method proposed by Breiman (1996). It uses bootstrap replicates (Efron & Tibshirani 1994) of the data in order to generate multiple versions of a classifier, here unpruned DTs as described in the previous section. The predictions of the trees are then aggregated by plurality votes. Fully grown decision trees can be considered as being very suited for this aggregation. They overfit the training data and guarantee thereby variance among the outcomes, which is in this case a desired feature.

RF builds upon the concept of BAG, but additionally incorporates the basic notion of random subspaces (Ho 1998). It differs in that the best split at each tree node is obtained from random subspaces with a defined dimensionality $<p$, for x^p. Here, the dimensionality of random subspaces *mtry* is set to $\lceil\sqrt{p}\rceil$. The number of trees, *ntree*, for BAG and RF is set to 500.

Support Vector Machines (SVM linear and SVM rbf)

The main notion of SVMs is to define the optimal hyperplane in the input space which separates data points of two classes, by convention $y \in (-1, 1)$. The optimization attempts to fix the empirical error on the training data and to maximize the margin between the hyperplane and the closest data points from each class. This follows the intuition that a fat margin decreases the risk of misclassifying unseen patterns (Vapnik 1998).

The separating hyperplane H is defined as $w \cdot x + b = 0$, where $x \in \mathbf{R}^p$ is any point on the plane, $w \in \mathbf{R}^p$ is the normal vector, i.e. in X orthogonal to the hyperplane, and $b \in \mathbb{R}$ is the bias.

The distance between the H and the coordinate system's origin can be expressed by $|b|/\|w\|$, with $\|w\|$ being the Euclidian norm of w. The data points of both classes closest to H are the so-called support vectors and their locations are defined to be on the hyperplanes $H_1 : w \cdot x + b = 1$ and $H_2 : w \cdot x + b = -1$, with H_1 and H_2 being parallel to H, hence sharing the same normal w. Their distances to the origin are given by $|1 - \mathrm{b}|/\|w\|$ and $|-1 - \mathrm{b}|/\|w\|$ respectively. The distances of H_1 and H_2 to H are equally $1/\|w\|$ and the margin, the distance between H_1 and H_2, is $2/\|w\|$.

The optimization problem now aims at finding H with the maximum margin by minimizing $\|w\|^2$, subject to the constraint that no data point lays in the margin area. This is termed *soft margin*. However, often the constraints are weakened to reduce the complexity and increase the generalization capacity of the solution. This is done by introducing a slack variable $\xi \geq 0$ that penalizes the misclassification of data points. The impact of ξ on the solution is controlled by the cost parameter C. The constrained optimization problem can be summarized as follows:

$$\mathit{minimize} \left[\frac{1}{2}\|w\|^2 + \frac{C}{n}\sum_{i=1}^{n}\xi_i \right] \tag{3}$$

$$\mathit{subject\ to} \begin{cases} w \cdot x + b \geq +1 - \xi_i \text{ if } y_i = +1 \\ w \cdot x + b \leq -1 - \xi_i \text{ if } y_i = -1 \end{cases} \tag{4}$$

The C parameter allows treating outliers, in remote sensing for example the spectral signal of a sparse material, and noise, e.g. mislabelled ground truth pixels. However, the still linear approach requires an adaptation unless the noise but the underlying processes, e.g. radiative transfer, cause problems because of the non-linearity of the system $X \rightarrow Y$. Then the input data can be mapped by a kernel function to a higher dimensionality, where it is more likely to follow a linear distribution that suits a hyperplane which represents a linear decision boundary. The mapping is defined by a kernel function $k(x_i, x_j) = \langle \phi(x_i) \,|\, \phi(x_j) \rangle$ which returns the dot product of two data points x_i and x_j, given a defined projection $\phi : X \rightarrow Z$ with the data points appearing only inside dot products with other points, this method allows to let the algorithms operate in X instead of Z space, which is commonly referred to as the *kernel trick* (Schölkopf & Smola 2002) because it saves computational costs. Often applied kernels are polynomial, sigmoid and Gaussian radial basis function (rbf). The rbf kernel is used in this study, which is defined as:

$$k\,(x_i, x_j) = \exp\,(-\,\sigma\,\|x_i - x_j\|^2),\ \sigma \in \mathbf{R}^+ \qquad (5)$$

with σ controlling the width of the Gaussian rbf. Both C and σ are parameters that have to be set by the user.

In this study, C and σ have been defined by a tuning grid (SVM rbf) or range (SVM linear) (Tab. 1). Furthermore, by default, all data is scaled prior to model generation to avoid that features with greater numeric range dominate those with smaller ranges. In order to extend from binary to multiclass cases with $c > 2$ classes, the *one-against-one* cascade is used, in which $c(c-1)/2$ binary models are trained to predict the final class by voting. For a comprehensive tutorial on support vector machines, which also outlines how to solve the constrained optimization problem (1)(2), the reader is referred to Burges (1998).

k-Nearest Neighbour (KNN)

KNN can be considered as a benchmark for more sophisticated approaches because it has a simple intuition and low computational demands. KNN does rather memorize than learn data. It relies on the closeness of still unclassified data to the k closest training data points in the input space. To measure closeness typically Euclidian distance is used. If $k > 1$, majority voting determines the final class.

2.3 *Object Feature Extraction*

Using eCognition software, a routine for object-based feature extraction has been developed which generates three segmentation levels from its input layers. Then, on each level, a defined set of object features layers (see Fig. 1 and Tab. 2 for examples) is computed and finally exported into a raster format for layer stacking and subsequent analysis. The set comprises features describing the objects' geometry and also those referring to the pixel

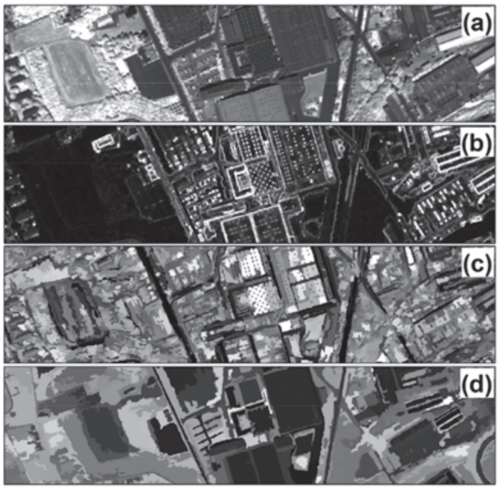

Fig. 1: Examples of spectral and spatial features: (a) CASI band #49 (831 nm) at pixel level; (b) *Mean difference to neighbours layer 1* (CASI band #11, 466 nm) at object level scale 20; (c) *Density* at object level scale 100; (d) *Standard deviation layer 4* (CASI band #49, 831 nm) at object level scale 500.

Tab. 2: Object features (see Trimble 2012 for details).

Features based on layer values (calculated per input layer)	Features based on object geometry
• Mean • Standard deviation • Mean difference to neighbours • GLCM (Gray Level Co-occurrence Matrix) homogeneity (all directions)	• Area • Border length • Length • Width • Density • Rectangular fit • Roundness • Shape index • Border index

values of the input layers to be calculated per input layer. The routine is applied for all images, whereas in the case of the hyperspectral datasets only four bands representing the blue, green, red and near-infrared range were considered as input layers in order to reduce data volume and redundancy. The number of object feature layers extracted per image can be obtained from Tab. 3.

To generate segmentation levels, multiresolution segmentation (Baatz & Schäpe 2000) was used. It is a region merging technique which has a scale parameter to control spectral and spatial heterogeneity constraints and thereby also the average objects' size. The definition of scale parameters was only based on a quick visual check without paying particular attention to the segmentation accuracy in terms of meaningful objects that represent geo-objects (Castilla & Hay 2008) which is a time-consuming process. The purpose of the visual check is ensuring that the fine level captures the detailed content of the scene while the coarse level should roughly represent the large structures in the image like crop fields and urban structure types. The finally selected scale parameters can be obtained from Tab. 3. Other segmentation parameters were kept as default (*shape*: 0.1; *compactness*: 0.5*;* equal *layer weights*).

2.4 *Evaluation Procedure*

With six algorithms (Tab. 1) and two input spaces, altogether twelve *algorithm-input space* constellations were empirically tested. Their performance was assessed within a nested cross-validation scheme which repeatedly generates training and test datasets from the available labelled set. This scheme enables the efficient use of typically sparse labelled data and ensures that both train and test sets come from the same distribution (Efron & Tibshirani 1994). The Cohen's kappa coefficient (κ) was used as accuracy measure (Congalton & Green 1999).

The nested approach (Fig. 2) was chosen because some algorithms require some test data for parameter tuning (Tab. 1) and these data instances should not be involved in the actual testing. Referring to Fig. 2, tuning parameters are tested with an inner cross-validation CV_i (5 cycles) that generates its CV_i *train sets* and CV_i *test sets* from the CV_o *train set*. Results of the *l* parameter setting candidates (Tab. 1) are compared and the best setting is used to parameterize a model for training on CV_o *train set* and testing on CV_o *test set* within the outer loop (10 cycles). For validating both the inner and the outer loop k-fold cross-validation was chosen. Only for the ROSIS dataset, where labelled data is plentiful (Tab. 3), 10-times random subsampling was chosen for the outer

Tab. 3: Overview of the experiments.

Case study		Object-based feature extraction		Input space (dimension)		Sample size (labelled pixels)
Mapping task	Image	Segmentation scale range	No. of object feature layers	*Spectral* (no. of bands)	*Spatial-spectral*	
Building detection	CASI	{20, 100, 500}	75 (25 per level)	72	147	2000
Building detection	Leica RC30	{10, 50, 250}	63 (21 per level)	3	66	2000
Mapping of pools, turf and tree/shrub vegetation	WV-2 Pan	{20, 100, 500}	39 (13 per level)	1	40	2059
Urban land cover classification	ROSIS	{20, 100, 500}	75 (25 per level)	102	177	99,114
Tree species classification	CASI	{20, 100, 500}	75 (25 per level)	72	147	400

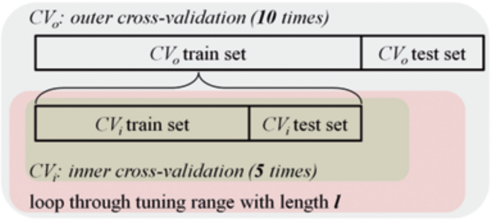
CV_o: outer cross-validation (10 times)
CV_o train set
CV_o test set
CV_i train set
CV_i test set
CV_i: inner cross-validation (5 times)
loop through tuning range with length l

Fig. 2: Nested cross-validation.

loop. This scheme draws in each cycle equally 50 training pixels per class, while the remaining part is for testing.

With respect to data variance, the classification models are trained and tested on the same partitions of the data by passing them in parallel through the nested process. Algorithms without tuning parameters simply skip the inner cross-validation.

3 Results and Discussion

An overview of the results is given by Fig. 3 which shows the distribution of κ coefficients obtained through the ten outer cross-validation cycles. Tab. 4 summarizes the cycles by listing the arithmetic mean. In addition, for each case study the best performing classification model was chosen for the presentation of error matrices and associated overall accuracies (OA) as well as producer's and user's accuracies (Tabs. 5–9). The error matrices were summed up from the outer crossvalidation cycles.

Comparing the results with respect to the input space shows that for all case studies almost all algorithms could benefit from the addition of spatial features. Given a particular algorithm, most improvements are quite distinct, indicated by an increased κ with non-overlapping interquartile ranges (Fig. 3). The less distinct cases with overlapping interquartile ranges concern DT, RF and KNN in conjunction with the CASI-based building detection case study. In contrast, no superiority of one input space over the other was obtained for KNN in conjunction with the WV-2 data which provides the mapping of pools, turf and tree/shrub vegetation. For KNN in conjunction with the RC30-based building detection, the spatial-spectral input space produced even significantly worse result. Overall, a relatively poor performance of the KNN in the spatial-spectral input spaces can be noticed. Possibly, this is because object-based features can have different value ranges which impacts on distance measuring in Euclidian space. In con-

Tab. 4: Cross-validation arithmetic mean (κ). The cell colours are defined by a linear red-to-green gradient which spans per case study from the minimum to the maximum κ value.

Input space	Algorithm	Arithmetic mean (κ)				
		CASI (buildings)	RC30	WV-2	ROSIS	CASI (trees)
Spectral	DT	0.407	0.210	0.559	0.838	0.416
	BAG	0.465	0.250	0.530	0.923	0.525
	RF	0.484	0.230	0.545	0.937	0.499
	SVM rbf	0.531	0.269	0.552	0.964	0.585
	SVM linear	0.420	0.219	0.534	0.964	0.543
	KNN	0.449	0.283	0.492	0.952	0.280
Spatial-spectral	DT	0.496	0.322	0.618	0.934	0.611
	BAG	0.556	0.446	0.719	0.971	0.738
	RF	0.565	0.452	0.717	0.974	0.766
	SVM rbf	0.604	0.535	0.700	0.992	0.703
	SVM linear	0.565	0.477	0.694	0.992	0.684
	KNN	0.478	0.147	0.477	0.970	0.580

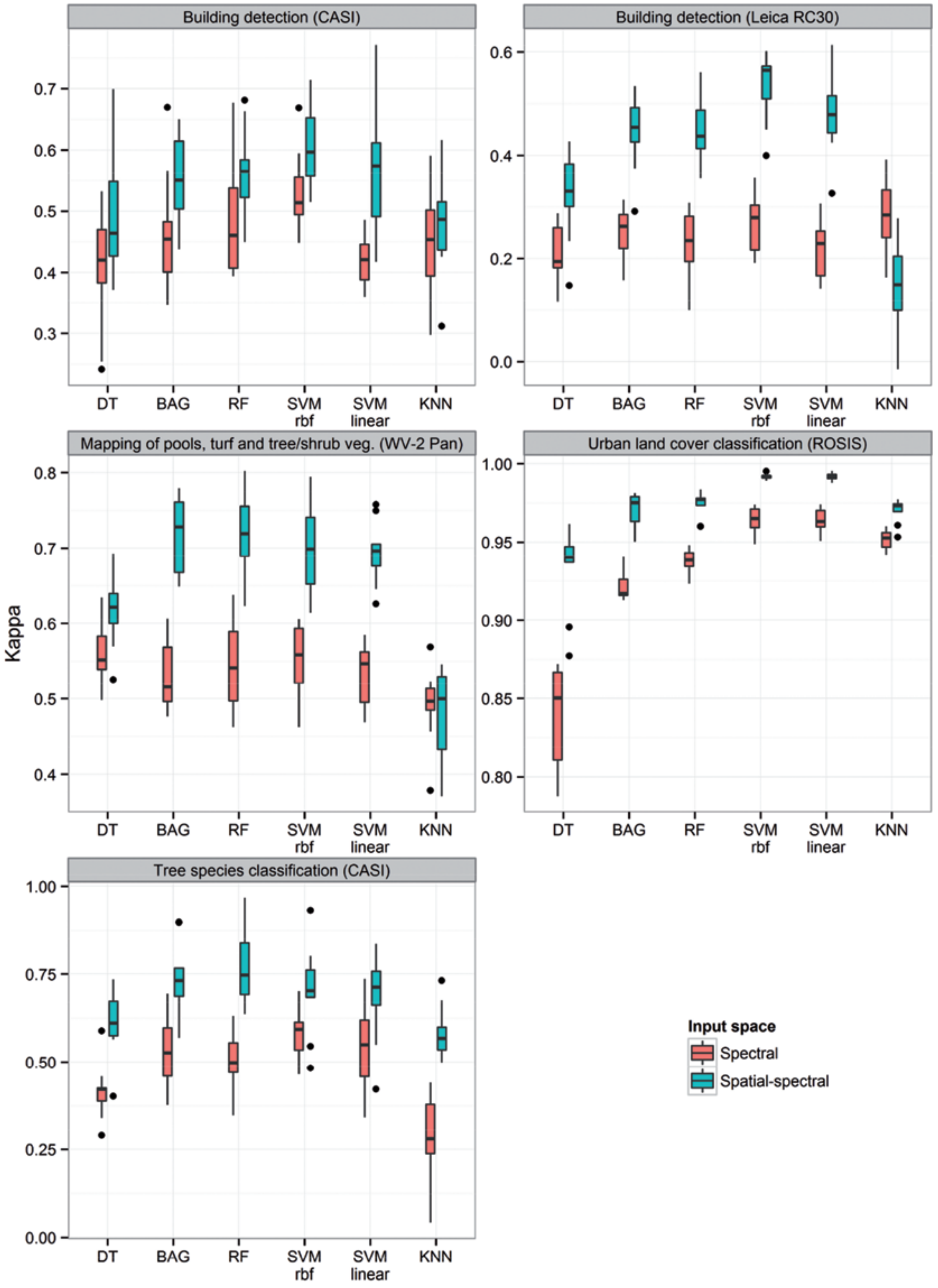

Fig. 3: Cross-validation boxplots (κ). The whiskers extend to the extreme data point which is no more than 1.5 times the interquartile range from the box. The points show outliers.

Tab. 5: Building detection with CASI data (SVM rbf).

Reference	**Roof**	**Background**
Prediction		
Roof	**311**	173
Background	94	**1422**
Producer	76.79	89.15
User	64.26	93.80
Overall	86.65	

Tab. 6: Building detection with RC30 data (SVM rbf).

Reference	**Roof**	**Background**
Prediction		
Roof	**236**	114
Background	169	**1481**
Producer	58.27	92.85
User	67.43	89.76
Overall	85.85	

trast, the results in the spectral input space were competitive with good results for both building detection and the ROSIS-based urban land cover classification case studies.

Comparing the performance of different algorithms in conjunction with the spatial-spectral input space, the results vary across the case studies (Tab. 4). For the CASI-based building detection, SVM rbf achieved the best result (κ 0.604, OA 86.65 %, Tab. 5), followed by SVM linear and RF. The results are encouraging if one takes into account that no additional elevation information, e.g. lidar or stereo models, was used and with respect to the difficult study area which includes heterogeneous, dense urban structures with different land uses and roof types. Possibly, a higher quantity and quality (small positional and thematic mismatch between the roof and the cadastre reference) of training data as well as a postprocessing of the results could improve the results.

For the RC30-based case study, again the SVM rbf algorithm yielded the best results (κ 0.535, OA 85.85 %, Tab. 6), followed by SVM linear and RF.

Other than for the building detection cases, the decision tree ensemble approaches (RF: κ 0.717, BAG: κ 0.719) outperformed the support vector machine approaches (SVM rbf: κ 0.700, SVM linear: κ 0.694) for the mapping of *pools*, *turf* and *trees/shrubs* with panchromatic WV-2 data. The error matrix (Tab. 7) shows high producer's and user's accuracies (> 83 %) for *pools* and *trees/shrubs*, but lower producer's accuracy (65.75 %) for *turf* which is in particular due to confusion with the *background* class. Possibly, paved areas like parking lots show similar geometrical and textural properties and cannot be differentiated by the limited spectral information. In this context, much improvement could be expected from the the inclusion of the eight multispectral bands of WV-2 (Wolf et al. 2012).

Tab. 7: Mapping of pools, turf and trees/shrubs with WV-2 data (BAG).

Reference	**Pool**	**Turf grass**	**Tree/ Shrub**	**Back-ground**
Prediction				
Pool	**36**	0	2	3
Turf grass	0	**144**	5	21
Tree/shrub	4	7	**684**	129
Background	2	68	102	**852**
Producer	85.71	65.75	86.25	84.78
User	87.80	84.71	83.01	83.20
Overall	83.34			

For the ROSIS-based application, the nine urban land cover classes have been best predicted by SVM rbf and SVM linear, achieving particular high agreement with the reference (both κ 0.992). However, true mapping accuracy can be expected to be much lower due to the non-random clustered distribution of the reference, which does not represent the image in its entirety. In comparison to the previous research on this benchmarking dataset (Palmason et al. 2005, Tarabalka et al. 2010), the obtained results show an improvement in terms of OA and producer's/user's accuracies (Tab. 8).

The classification of urban tree species using hyperspectral data was best achieved by RF (κ 0.766, OA 82.25 %). The producer's accuracies range from 72 % (*Acer platanoides*) to 100 % (*Tilia tomentosa*) and the user's accuracies from 41 % (*Crataegus x lavallei*) to 92 % (*Acer pseudoplatanus*) (Tab. 9).

Summarizing the algorithm performances, it can be noted that support vector machine as well as tree ensemble approaches provided reliable solutions even for quite diverse mapping scenarios (in terms of sensors, classes and labelled information), whereas decision tree and nearest neighbour were not competitive. This strengthens the statements of previous research studies reporting on good performances of support vector machine and random forest approaches for high dimensional problems (Duro et al. 2012, Huang et al. 2002, Waske et al. 2009). Comparing both tree ensemble methods, RF achieved in four of five case studies the higher κ. The difference is only marginal, but aligns with Breiman (2001) who states the theoretical and empirical superiority of RF. Among the support vector machine kernels, rbf provided better results in four of five case studies.

Tab. 8: Urban land cover classification with ROSIS data (SVM rbf).

Reference	**Water**	**Trees**	**Asphalt**	**Bricks**	**Bitumen**	**Tiles**	**Shadows**	**Meadows**	**Soil**
Prediction									
Water	**465272**	0	0	0	0	0	0	0	0
Trees	0	**35560**	302	0	2	0	0	5	2
Asphalt	0	1848	**12944**	0	122	0	0	0	0
Bricks	0	0	0	**25320**	71	256	23	289	0
Bitumen	0	0	17	6	**49710**	57	0	15	0
Tiles	0	0	0	12	0	**57935**	0	827	0
Shadows	414	0	0	23	3	17	**34266**	14	0
Meadows	0	0	27	19	2	15	0	**282211**	0
Soil	14	2	0	0	0	0	1	19	18998
Producer	99.91	95.05	97.40	99.76	99.60	99.41	99.93	99.59	99.99
User	100.00	99.14	86.79	97.54	99.81	98.57	98.64	99.98	99.88
Overall	99.55								

Tab. 9: Urban tree species classification with CASI data (RF).

Reference	**Ac. pl.**	**Ac. ps.**	**Ac. sa.**	**Ae. ca.**	**Be. pe.**	**Cr. la.**	**Pl. ac.**	**Ti. to.**
Prediction								
Acer platanoides	**18**	10	0	0	0	0	0	0
Acer pseudoplatanus	5	**159**	2	1	3	1	0	0
Acer saccharinum	0	6	**36**	0	0	0	1	0
Aesculus carnea	0	9	0	**24**	0	0	0	0
Betula pendula	0	7	0	0	**22**	0	0	0
Crataegus lavallei	0	10	0	0	0	**7**	0	0
Platanus acererifolia	2	8	0	0	1	0	**29**	0
Tilia tomentosa	0	3	1	0	1	0	0	**34**
Producer	72.00	75.00	92.31	96.00	81.48	87.50	96.67	100.00
User	64.29	92.98	83.72	72.73	75.86	41.18	72.50	87.18
Overall	82.25							

4 Summary and Conclusion

Comparing different machine learning algorithms, the value of adding objects-based features into per-pixel classification was investigated using different urban remote sensing case studies. Based on the experimental results it can be concluded that the addition of these features can improve the classification performance if combined with an appropriate learning algorithm. Especially support vector machines and decision tree ensemble approaches performed well in this context. Moreover, this work presents a framework in which object-based features are accessed on the pixel domain. This can be an advantage for the application of machine learning with respect to unbiased data representation. Moreover, it can be taken into account as an alternative, straightforward option if an integration of the pixel and the object domain is intended, e.g. in case of Tarabalka et al. (2009) and Wang et al. (2004).

Acknowledgement

The author would like to thank Prof. Paolo Gamba of the University of Pavia, Italy, for providing the ROSIS image data along with the labelled reference information.

References

Baatz, M. & Schäpe, A., 1999: Object-Oriented and Multi-Scale Image Analysis in Semantic Networks. – 2nd International Symposium: Operationalization of Remote Sensing, 16–20 August, ITC, Niederlande.

Baatz, M. & Schäpe, A., 2000: Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation. – Angewandte Geographische Informationsverarbeitung **XII**: 12–23.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. & Heynen, M., 2004: Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. – ISPRS Journal of Photogrammetry and Remote Sensing **58** (3–4): 239–258.

Blaschke, T., Lang, S. & Hay, G.J., 2008: Object-based image analysis. – 817 p., Springer, Berlin.

Blaschke, T., 2010: Object based image analysis for remote sensing. – ISPRS Journal of Photogrammetry and Remote Sensing **65** (1): 2–16.

Breiman, L., Friedmann, J.H., Olshen, R.A. & Stone, C.J., 1984: Classification and regression trees. – Chapman & Hall, New York, USA.

Breiman, L., 1996: Bagging predictors. – Machine Learning **24** (2): 123–140, New York, USA.

Breiman, L., 2001: Random Forests. – Machine Learning **45** (1): 5–32.

Burges, C.J., 1998: A Tutorial on Support Vector Machines for Pattern Recognition. – Data Mining and Knowledge Discovery **2** (2): 121–167.

Burnett, C. & Blaschke, T., 2003: A multi-scale segmentation/object relationship modelling methodology for landscape analysis. – Ecological Modelling **168** (3): 233–249.

Castilla, G. & Hay, G.J., 2008: Image objects and geographic objects. – Blaschke, T., Lang, S. & Hay, G.J. (eds.): Object-Based Image Analysis. Spatial Concepts for Knowledge-Driven Remote Sensing Applications: 91–110, Springer-Verlag, Berlin.

Congalton, R.G. & Green, K., 1999: Assessing the accuracy of remotely sensed data. – Lewis Publications, Boca Raton, USA.

Duro, D.C., Franklin, S.E. & Dubé, M.G., 2012: A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. – Remote Sensing of Environment **118:** 259–272.

Efron, B. & Tibshirani, R., 1994: An introduction to the bootstrap. – 436 p., Chapman & Hall, New York, USA.

Herold, M., Gardner, M. & Roberts, D.A., 2003: Spectral resolution requirements for mapping urban areas. – IEEE Transactions on Geoscience and Remote Sensing **41** (9): 1907–1919.

Hof, A. & Schmitt, T., 2011: Urban and tourist land use patterns and water consumption: Evidence from Mallorca, Balearic Islands. – Land Use Policy **28** (4): 792–804.

Ho, T.K., 1998: The random subspace method for constructing decision forests. – IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (8): 832–844.

Huang, C., Davis, L.S. & Townshend, J.R.G., 2002: An assessment of support vector machines for land cover classification. – International Journal of Remote Sensing **23** (4): 725–749.

Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A., 2004: kernlab – An S4 Package for Kernel Methods in R. – Journal of Statistical Software **11** (9): 1–20.

Lu, D. & Weng, Q., 2007: A survey of image classification methods and techniques for improving classification performance. – International Journal of Remote Sensing **28** (5): 823–870.

Novack, T., Esch, T., Kux, H.J.H. & Stilla, U., 2011: Machine Learning Comparison between WorldView-2 and QuickBird-2-Simulated Imagery Regarding Object-Based Urban Land Cover Classification. – Remote Sensing **3** (10): 2263–2282.

Palmason, J., Benediktsson, J.A., Sveinsson, J. & Chanussot, J., 2005: Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis. – IEEE International Geoscience and Remote Sensing Symposium: 176–179.

R Core Team, 2012: R: A Language and Environment for Statistical Computing. – http://www.R-project.org (4.11.2012).

Schölkopf, B. & Smola, A.J., 2002: Learning with kernels. – MIT Press, Cambridge, Massachusetts, USA.

Stehman, S.V. & Czaplewski, R.L., 1998: Design and Analysis for Thematic Map Accuracy Assessment. – Remote Sensing of Environment **64** (3): 331–344.

Svetnik, V., Liaw, A., Tong, C., Culberson, J., Sheridan, R. & Feuston, B., 2003: Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. – Journal of Chemical Information and Modeling **43** (6): 1947–1958.

Tarabalka, Y., Benediktsson, J.A. & Chanussot, J., 2009: Spectral-Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques. – IEEE Transactions on Geoscience and Remote Sensing **47** (8): 2973–2987.

Tarabalka, Y., Fauvel, M., Chanussot, J. & Benediktsson, J.A., 2010: SVM- and MRF-Based Method for Accurate Classification of Hyperspectral Images. – IEEE Geoscience and Remote Sensing Letters **7** (4): 736–740.

Therneau, T.M. & Atkinson, E.J., 1997: An introduction to recursive partitioning using the rpart routines. – http://mayoresearch.mayo.edu/mayo/research/biostat/upload/61.pdf (13.11.2012).

Trimble, 2012: eCognition® Developer 8.8. Reference Book. – Trimble Germany GmbH.

Vapnik, V.N., 1998: Statistical learning theory. – Wiley, New York, USA.

Venables, W.N. & Ripley, B.D., 2002: Modern Applied Statistics with S. – **4th** ed., Springer, New York, USA.

Wang, L., Sousa, W.P. & Gong, P., 2004: Integration of object-based and pixel-based classification for mapping mangroves with IKONOS imagery. – International Journal of Remote Sensing **25** (24): 5655–5668.

Waske, B., Benediktsson, J.A., Árnason, K. & Sveinsson, J.R., 2009: Mapping of hyperspectral AVIRIS data using machine-learning algorithms. – Canadian Journal of Remote Sensing **35** (S1): 106–116.

Wolf, N., Hof, A. & Jürgens, C., 2012: Machine learning for the exhaustive evaluation of object-based feature spaces. – GEOBIA **2012:** 267–272.

Address of the Author:

M.Sc. Geography Nils Wolf, Ruhr-Universität Bochum, Geographisches Institut, D-44801 Bochum, Germany, Tel.: +49-234-32-23380, Fax: +49-234-32-14180, e-mail: nils.wolf@rub.de

Manuskript eingereicht: November 2012
Angenommen: März 2013