Object-based Post-classification Relearning

Christian Geiß, *Member, IEEE*, and Hannes Taubenböck

German Aerospace Center (DLR), German Remote Sensing Data Center (DFD),
Münchner Straße 20, 82234 Oberpfaffenhofen-Weßling, Germany
Tel.: +49 8153 28 1255, fax: +49 8153 28 1445
christian.geiss@dlr.de; hannes.taubenboeck@dlr.de

***Abstract*— In this letter, we present an object-based post-classification relearning approach for enhanced supervised remote sensing image classification. Conventional post-classification processing techniques aim to enhance the classification accuracy by imposing smoothness priors in the image domain (based on e.g., majority filtering or markov random fields). In contrast to that, here, a supervised classification model is learned for a second time with additional information generated from the initial classification outcome to enhance discriminative properties of relearned decision functions. This idea is followed within an object-based image analysis framework. Therefore, we model spatial-hierarchical context relations with the preliminary classification outcome by computing class-related features using a triplet of hierarchical segmentation levels. Those features are used to enlarge the initial feature space and impose spatial regularization in the relearned model. We evaluate the relevance of the method in the context of classifying of a high resolution multispectral image, which was acquired over an urban environment. Experimental results show enhanced classification accuracy using this method compared to both a per-pixel based approach and outcomes obtained with a conventional object-based post-classification processing technique (i.e., object based voting).**

***Index Terms*—object-based image analysis, relearning, classification postprocessing, SVM**

I. INTRODUCTION

Methods for derivation of thematic information from remote sensing data have been a major research subject over the past decades. Thereby, supervised classification approaches are very popular due to their accuracy, robustness, and flexibility [1]. The idea of such methods is to infer a rule (e.g., a decision function) from limited but properly encoded prior knowledge to assign a class label to unseen instances of the domain under analysis. Thereby, the concept of spectral per-pixel classification has been extended especially for remote sensing data with a ground sampling distance considerably higher than the objects of interest. This kind of data became in particular available with the advent of sensors such as IKONOS, QuickBird, WorldView I-III, or GeoEye, among others. The high geometric resolution can induce high intraclass and low interclass variabilities especially in heterogeneous environments such as urban areas. This causes dominantly the well-known salt-and-pepper effect in the classification outcome, when relying solely on individual spectral signatures of pixels [2].

To cope with this problem, three distinguishable strategies were followed in the past: (i) Deployment of features that aim to incorporate spatial relations before classification (referred to as spectral-spatial classification). Those kinds of features take the neighborhood of individual pixels into account such as morphological operators (e.g, [3]-[5]), or texture measures (e.g.,[6]-[8]); (ii) Partition of the image with a segmentation algorithm into objects and usage of e.g., the spectral means and spatial-hierarchical context characteristics for classification on segment level (referred to as object-based image analysis (OBIA)) [2], [9], [10]; (iii) The refinement of the classification outcome by classification postprocessing (CPP) [11]. With respect to the latter, the majority of approaches aim to refine the initial classification outcome by taking advantage of spatial occurrence and alignment of class labels and eventually *relabel them in the image domain*, based on e.g., majority filtering [12], or markov random fields [13].

Recently, Huang et al. [11] proposed two CPP strategies that deploy the concept of *relearning*. Thereby, a supervised classification model is learned for a second time with additional features derived from the initial classification outcome. In particular, they propose to compute a primitive co-occurrence matrix and local class histograms for characterization of the spatial occurrence and alignment of class labels in the feature space. Experimental results showed better accuracies compared to a per-pixel approach and traditional CPP methods.

In this letter, we adapt the idea of relearning and extend it in the context of an OBIA approach, referred to as object-based post-classification relearning (OBR). We make use of the unique capabilities of OBIA, which allow addressing a multiplicity of spatial scales and hierarchies within the image domain [9]. To this purpose, the initial classification outcome is aggregated first on a triplet of segmentation levels, which were generated from the remote sensing imagery. The information is used to compute multiple class-related features, which are subsequently used for relearning the model. This is intended to allow for representing complex spatial-hierarchical context relations of class labels in the feature space in order to simultaneously enhance discriminative properties of relearned decision functions and impose adequate smoothness priors. To demonstrate the relevance of the method we compare it to the current most popular OBIA CPP strategy, named object-based voting (OBV) [11]. Additionally, comparisons to a per-pixel based approach are made based on both spectral and

spectral-spatial feature sets.

The remainder of the letter is organized as follows: section II details the proposed OBR method. We describe the experimental setup in section III and report results of actual experiments in section IV. Concluding remarks are given in section V.

II. Proposed Methodology

The OBR approach consist of three consecutive main steps (Fig 1): i) an initial supervised model is learned and applied to the image; ii) the labeled pixels are aggregated with respect to a triplet of segmentation levels and class-related features are generated from them; iii) The model is relearned with the additional features and finally applied to the image.

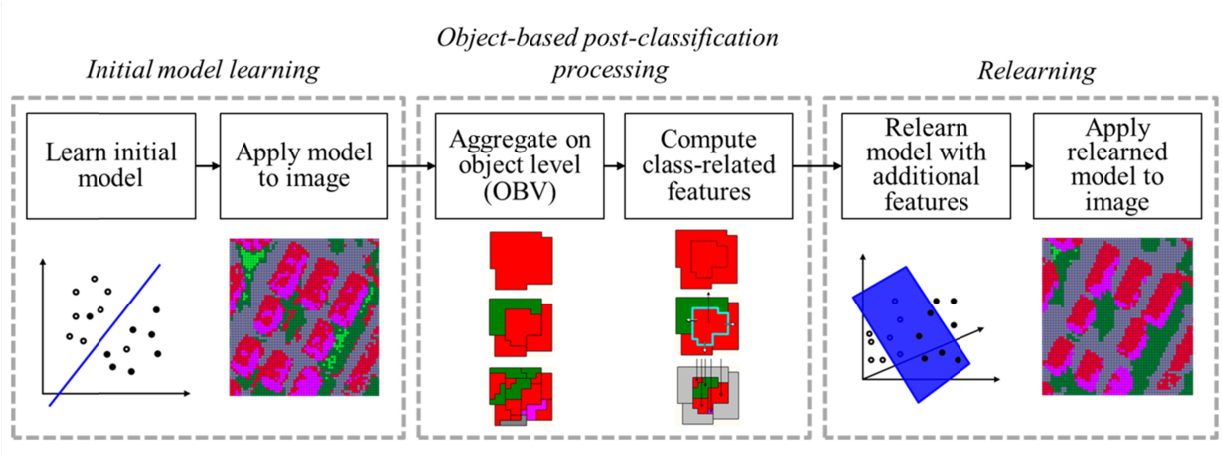

Fig. 1. Scheme of the proposed object-based post-classification relearning approach.

Let us consider an image (e.g., acquired by a multi- or hyperspectral sensor) I_w constituted by multiple spectral bands F_w. The spectral information allows deriving further features F_e by e.g., exploiting both spectral and spectral-spatial properties. As such, $I_{we} \in \mathbb{R}^{(F_w+F_e)}$ represents the composed image when stacking F_w and F_e.

Following a supervised approach, the training set $\mathrm{S} = \{\mathrm{X}, \mathrm{Y}\}$ is given, where $\mathrm{X} = \{\mathbf{x}_l\}_{l=1}^{n} \in I_{we}$ is a subset of I_{we} made up of n labeled samples, and associated class labels $\mathrm{Y} = \{y_l\}_{l=1}^{n} \in \{1, \ldots, q, \ldots, Q\}$. Then, a classification approach is deployed to provide a class label for all remaining unlabeled instances $X^* = \{\mathbf{x}_u^*\}_{u=1}^{m} \in I_{we}$. In this manuscript, we consider the soft margin SVM (C-SVM) [14] for solving the classification problem. There, the learned decision function for a two-class problem allows assigning a class label to an instance of unknown class membership $\mathbf{x}_u^*$:

$$f(\mathbf{x}_u^*) = \mathrm{sign}\left(\sum_{l=1}^{\mathrm{n}} y_l \alpha_l K(\mathbf{x}_l, \mathbf{x}_u^*) + b\right)$$

with $K(\mathbf{x}_l, \mathbf{x}_u^*)$ being the kernel matrix, which quantifies the similarity between the test vector $\mathbf{x}_u^*$ and the l-th support vector, and α are the support vector coefficients with affiliated class labels $y_l \in \{-1, +1\}$. Only samples that become a support vector are needed to define the

model, which allows for building robust models with a high generalization capability based on a comparatively small number of labeled training samples. For solving a multi-class problem an individual model can be learned based a one-against-all scheme and the test vector is assigned to the class, which minimizes the class-specific decision function [15].

In parallel, I_w is subject to a multilevel, hierarchical segmentation procedure [9], [10], [16]. Thereby, I_w is partitioned with an arbitrary segmentation algorithm at a generic segmentation level s in N^s objects O_i^s $(i = 1, 2, \ldots, N^s)$. To establish an unambiguous hierarchy of segmentation levels, the following constraint must be fulfilled:

$$\bigcup_{o_i^{s-1} \subseteq o_j^s} O_i^{s-1} = O_j^s$$

This relation ensures that an object at segmentation level $s-1$ must be included in only one object at level s [10]. The subsequent computation of features is based on a triplet of generated hierarchical segmentation levels $\mathrm{M} \in \{s-1, s, s+1\}$. This is done to be able to encode both (topological) neighborhood relationships and hierarchical relationships, which are in general important constituent properties of OBIA techniques [9] (Fig. 2).

For aggregation of the class labels obtained with equation 1 to M, we deploy a crisp OBV scheme [11], what corresponds to a majority vote per object:

$$Q(O) = \arg\max_{q \in Q} \left(\frac{1}{N^O} \sum_{v \in O} \tau(Q(v) = q) \right)$$

with Q being the labeling space, $Q(O)$ the final label of object O, τ is an indicator function capturing the number of times that the labeled pixels v within an object feature class label q, and N^o is the number of pixels of an object.

Based on this information, we compute a set of class-related features F_q from the initial classification outcome (Fig. 2). From a reference object O^s the relative areas of classes of sub-objects A_q^{s-1} are quantified:

$$A_q^{s-1} = \frac{1}{N^O} \sum_{v \in O} v_q^{s-1}; \quad A_q^{s-1} \in \left\{ \mathbb{R}_+ \middle| 0 \le A_q^{s-1} \le 1 \right\}$$

with v_q^{s-1} being the labeled pixels of sub-objects of O^s, which feature class label q. The relative borders to classes of neighbor objects P_q^s are computed as follows:

$$P_q^s = \frac{1}{D^O} \sum d_q^s; \quad P_q^s \in \left\{ \mathbb{R}_+ \middle| 0 \le P_q^s \le 1 \right\}$$

with d_q^s being the labeled pixels adjacent to O^s, which feature class label q, and D^O is the number of border pixel of O^s. In addition, the class assignments at segmentation level $Q(O^s)$,

and $s+1$ $Q(O^{s+1})$ are taken into account. To represent those categorical features adequately for the C-SVM, the representation of the individual class assignments is implemented according to a 1-of-K coding scheme.

Finally, F_q is stacked to I_{we}, i.e, so that $I_{weq} \in \mathbb{R}^{(F_w+F_e+F_q)}$ represents the new image. Equation 1 is solved again with the extended feature space to provide the final classification solution.

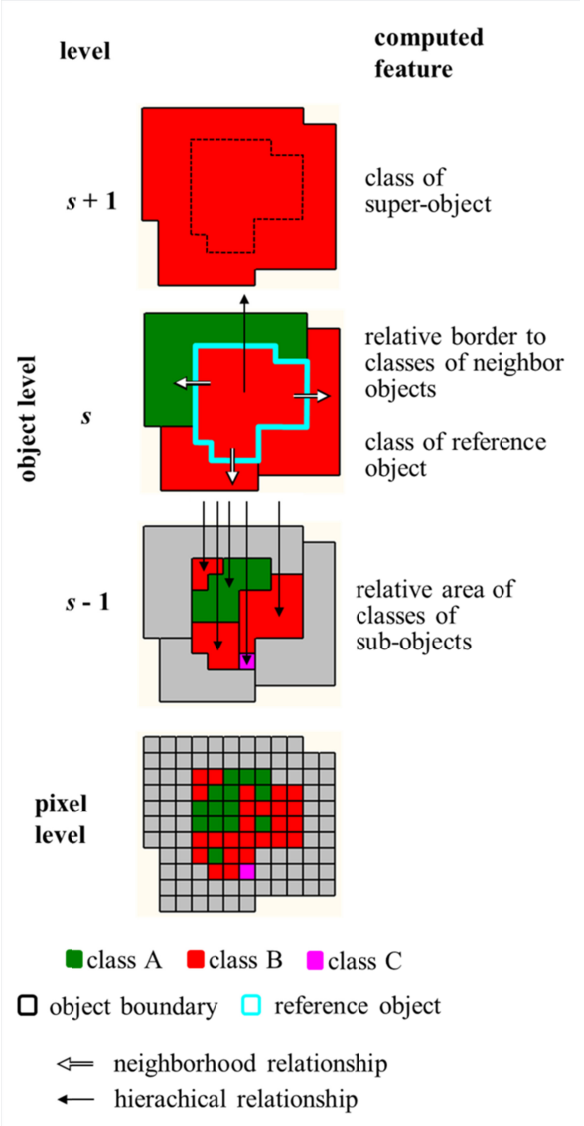

Fig. 2. Computation of class-related features with a triplet of hierarchical segmentation levels to incorporate (topological) neighborhood relationships and hierarchical relationships.

III. Experimental Setup

The data set considered in the experiments is a multispectral image of 2000 × 2000 pixels acquired over the city of Cologne from the QuickBird sensor with a spatial resolution of 0.65 m (Fig. 3a). However, we resampled the image to a spatial resolution of 2 m with a nearest neighbor interpolation to reduce the computational burden for the experiments.

In experiment I (sec. IV A) we use the four spectral bands for classification in conjunction with the normalized differenced vegetation index and two brightness-related features. In experiment II (sec. IV B) we incorporate a set of morphological operators (i.e., erosion, dilation, opening, closing, opening by top hat, and closing by top hat) [5] on the panchromatic band with linear ascending sizes of the square-shaped structuring element $B = \{3, 5, \dots, 13\}$ for spectral-spatial classification. Labeled samples (Fig. 3b) have been split into two spatially disjunct subsets. In both experiments a varying number of labeled samples have been drawn randomly from the training data pool (Fig. 3c) and generalization capabilities of the learned models are estimated with the testing data (Fig. 3d).

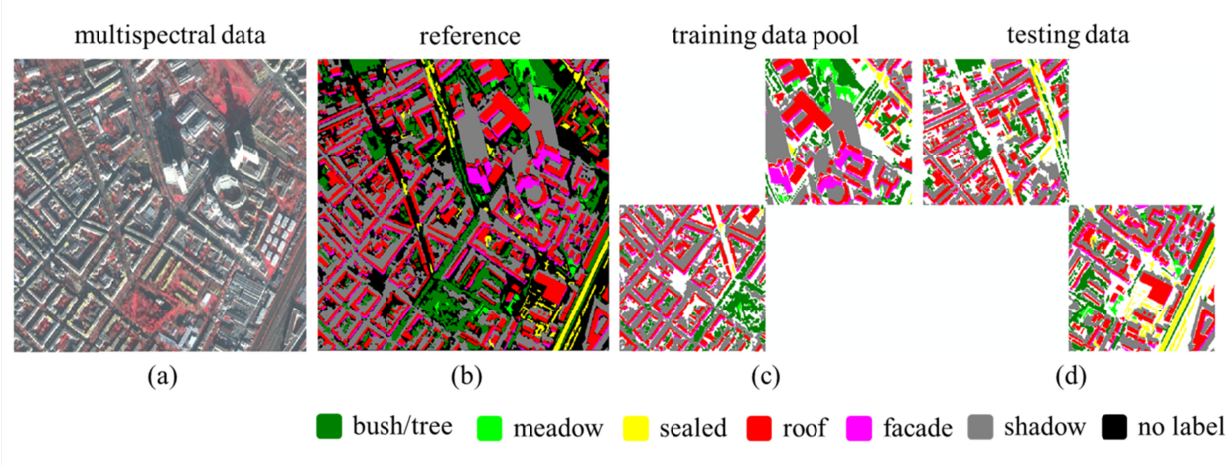

Fig. 3. Experimental data; (a) multispectral image from QuickBird acquired over the city of Cologne (Germany); (b) reference pixels available; (c) pool of labeled pixels used for stratified random selection and subsequent model learning; (d) labeled pixels used for validation.

We use a bottom-up region-growing segmentation algorithm for partition of I_w (i.e., fractal net evolution approach as implemented in the software environment eCognition [17]). In the experiments we put more emphasize on shape heterogeneity rather than on grey-value heterogeneity with respect to the segmentation algorithm. This is due to the fact that man-made structures such as buildings and other elements of urban environments have distinct shape and size properties, unlike e.g., natural features. Analogously, the weights for heterogeneity of smoothness and compactness were maintained equal (i.e., shape: 0.7, color: 0.5) and kept constant. The scale parameters for M, which determine - briefly speaking - the size of the objects, were chosen in a way that i) at $s - 1$ majority of real world objects are represented by several segments (i.e., over-segmentation), ii) at s majority of real world objects are represented by a single affiliated segment, and iii) at $s + 1$ majority of real world objects are represented in conjunction with other real world objects by a single segment (i.e., under-segmentation).

For the *C*-SVM, we deployed Gaussian RBF kernels, that take the form $K(\mathbf{x}_l, \mathbf{x}_u^*) = \exp(-\|\mathbf{x}_l - \mathbf{x}_u^*\|^2 / 2\sigma^2)$. Learning the most appropriate *C*-SVM in conjunction with a RBF kernel requires the definition of the cost-parameter C and the kernel-width parameter γ. We carried out an exhaustive optimization of hyperparameters $\Phi \in \{C, \gamma\}$: $C = \{2^{-4}, 2^{-3}, \dots, 2^{12}\}$ and $\gamma = \{2^{-5}, 2^{-4}, \dots, 2^{3}\}$. It can be noted that the same M was deployed for OBV and the computation of the features for OBR to ensure *ceteris paribus*-like conditions

for comparison of the methods. In this sense, also optimization of Φ for the *C*-SVM was not subject to a simultaneous optimization but a consecutive one with respect to OBV and OBR: first an optimal value for C and γ is determined and the affiliated (per-pixel) classification outcome is consistently used for subsequent processing. Alternatively, one could optimize Φ simultaneously for the initial classification in conjunction with the OBV and OBR approach. However, this would induce a considerable higher computational burden and more significantly would bias here the comparability of the different methods, since results would be compared that may rely on different initial classification outcomes.

IV. Experimental Results

A. Spectral classification

Estimated κ statistics obtained from different model runs with varying number of labeled samples included are revealed in Fig. 4 (It can be noted that we report only the best accuracy obtained from M for the OBV approach). From there, it can be seen that the proposed technique achieves the highest mean κ statistics over all runs. It is followed by OBV, whereas the per-pixel approach performs least favorable: e.g., given 1000 labeled samples for model training, the per-pixel approach reaches a mean κ statistic of 0.696 (±0.005), 0.698 (±0.005) for OBV, and 0.711 (±0.006) for OBR. Reasonably, we observe in this example that classification results obtained with a fewer number of labeled training samples are less accurate and stable. Thereby, also differences in methods become more marginal. Especially with very few samples (per class), problems associated with the "Hughes phenomenon" (which states that for a limited amount of samples the predictive power decreases as the dimensionality of the feature vector increases) can occur (note that OBR operates with an increased number of features).

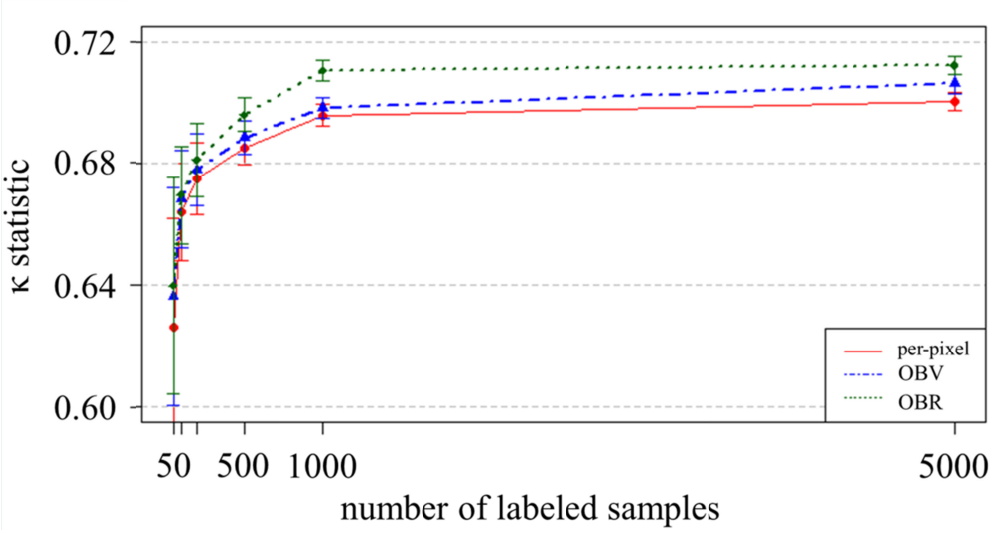

Fig. 4. Mean and standard deviation of κ statistic for spectral classification of the different approaches based on multiple model runs with varying number of labeled samples included.

Nevertheless, when inspecting the respective classification outcome, remarkable differences appear (Fig. 5). It can be seen that the per-pixel approach features the typical salt-and-pepper

effect in the classification outcome. Instead, both OBV and OBR provide solutions, which are spatially more homogeneous. Thereby, OBR is able to ensure a spatially regularized solution with an increased accuracy compared to OBV. This confirms the capability of the proposed methodology to deploy the intrinsic patterns in the initial classification outcome for learning a model with enhanced discriminative properties and adequate smoothness priors.

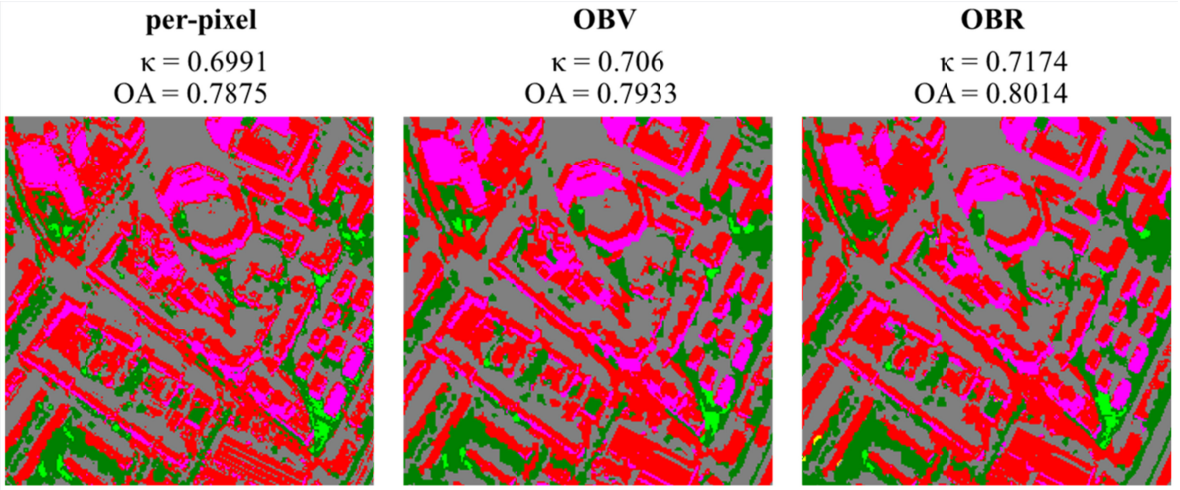

Fig. 5. Selected results for spectral classification obtained with 1000 labeled pixels for a per-pixel classification, OBV, and OBR.

B. Spectral-spatial classification

To get further insights on the performance of the methods when also spatial information is included, we classify the imagery under consideration of features which take the neighborhood of individual pixels into account (morphological operators). Estimated κ statistics and overall accuracies (OA) obtained from repeated model runs with varying samples included are revealed in Table I (note that we have enlarged the number of labeled samples compared to the previous experiment to avoid problems associated with the "Hughes phenomenon", since feature vectors have considerably increased in this experiment). In accordance with the previous experiment, OBR features consistently the highest classification accuracies. Notably, it is followed by the per-pixel approach, and here, OBV performs worst.

TABLE II MEAN (μ) AND STANDARD DEVIATION (σ) OF κ STATISTIC AND OA OBTAINED FROM 20 REALIZATIONS WITH VARYING CONFIGURATIONS OF 10,000 LABELED SAMPLES

approach	κ	OA
	μ(σ)	μ(σ)
per-pixel	0.7166 (0.0032)	0.7999 (0.0023)
OBV	0.7085 (0.0033)	0.7947 (0.0025)
OBR	**0.7411** (0.0055)	**0.8161** (0.0041)

This result becomes comprehensible, when inspecting the classification outcome (Fig. 6). First, it can be noted that the spatial features allow reducing the salt-and-pepper effect for the per-pixel approach, whereby classification accuracy is simultaneously increased. Instead, OBV tends to overgeneralize the initial classification outcome (although real world objects of

interest of reported results were dominantly represented by a multiplicity of corresponding segments). This characteristic becomes evident when focusing on the class “sealed” (depicted in yellow). Generally, it was very challenging to separate this class especially from the class “roof” in this example due to the spectral similarity of the classes and comparatively coarse spectral resolution of the data. In experiment I, vast parts of sealed surfaces were allocated incorrectly to the class “roofs”. In experiment II, the spatial information allows some improvements with respect to separability. However, sealed areas, as identified by the per-pixel approach, are largely “absorbed” by the OBV approach, what induces a decrease of accuracy. Instead, OBR features the unique capability to identify relevant shares in this setting correctly, what constitutes a distinctive level of accuracy.

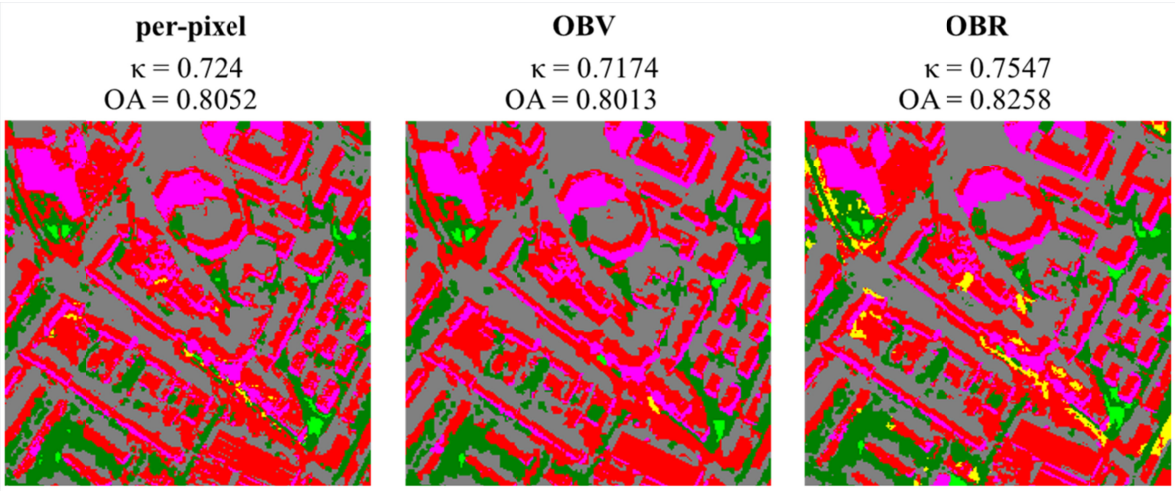

Fig. 6. Selected results for spectral-spatial classification obtained with 1000 labeled pixels for a per-pixel classification, OBV, and OBR.

V. Conclusions and Outlook

In this letter, we have proposed a novel post-classification relearning strategy within the OBIA framework to enhance classification accuracy. As such, this study extends recently postulated ideas of post-classification relearning. Notably, our results underline the findings of Huang et al. [11], which suggest that the general strategy of *relearning* can be superior compared to traditional CPP methods, which *relabel classification outcomes in the image domain*. This superordinate observation is suggested to be subject to a further scientific contemplation in future work.

More specifically, with respect to this study, future work may include the computation of additional spatial-hierarchical features from the initial classification outcome (e.g., utilize a larger number of segmentation levels or describe the spatial assemblage of class labels of sub-objects with indices to characterize categorical map patterns). Moreover, only the most relevant features based on feature selection or dimensionality reduction strategies can be considered for model learning. Thereby, models can also be learned iteratively with multiple initial classification outcomes and affiliated OBR strategy. Thereby, it should be evaluated if such a post-classification technique can also further enhance classification accuracy in the context of complex multilevel classification approaches (e.g., [10], [18]).

VI. ACKNOWLEDGEMENTS

This work was supported by the German Federal Ministry for Economic Affairs and Energy's initiative "Smart Data – innovations from data" under grant agreement: "smart data for catastrophe management (sd-kama)". We thank the anonymous reviewers for helpful comments on the initial manuscript.

VII. REFERENCES

[1] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*. Hoboken, NJ: Wiley, 2009.
[2] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 2-16, Jan. 2010.
[3] R. Haralick, S. Sternberg, and X. Zhuang "Image analysis using mathematical morphology," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-9, no. 4, pp. 532–550, Jul. 1987.
[4] P. Soille, Morphological Image Analysis, 2nd ed. Berlin, Germany: Springer-Verlag, 2004.
[5] D. Tuia, F. Pacifici, M. Kanevsi, W.J. Emery, "Classification of Very High Spatial Resolution Imagery Using Mathematical Morphology and Support Vector Machines," IEEE Trans. Geosci. Remote Sens., vol. 47, no. 11, pp. 3866 - 3879, Nov. 2009
[6] R. Haralick, "Statistical and Structural Approaches to Texture," Proc. IEEE, vol. 67, no. 5, pp. 786-804, May 1979.
[7] F. Pacifici, M. Chine, W.J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," Remote Sensing of Environment, vol. 113, no. 6, pp.1276-1292, Jun 2009.
[8] M. Musci, R. Queiroz Feitosa, G.A.O.P. Costa, M.L. Fernandes Velloso, "Assessment of Binary Coding Techniques for Texture Characterization in Remote Sensing Imagery," IEEE Geosci. Remote Sens. Letters, vol. 10, no. 6, pp. 1607-1611, Oct. 2013
[9] T. Blaschke et al., "Geographic object-based image analysis: A new paradigm in remote sensing and geographic information science," ISPRS Int. J. Photogramm. Remote Sens., vol. 87, no. 1, pp. 180–191, 2014.
[10] 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.
[11] X. Huang, Q. Lu, L. Zhang, and A. Plaza, "New postprocessing methods for remote sensing image classification: A systematic study," IEEE Trans. Geosci. Remote Sens., vol. 52, no. 11, pp. 7140–7159, Nov. 2014
[12] J. Stuckens, P. R. Coppin, and M. E. Bauer, "Integrating contextual information with per-pixel classification for improved land cover classification," Remote Sens. Environ., vol. 71, no. 3, pp. 282–296, Mar. 2000
[13] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
[14] C. Cortes, V. Vapnik, "Support Vector Networks," *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.
[15] F. Melgani, L. Bruzzone "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines," IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778 - 1790, Jul. 2004.
[16] Pal R, Pal K (1993) A review on image segmentation techniques. Pattern Recognition, 26 (9):1277-1294
[17] M. Baatz, A. Schäpe (2000). Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In J. Strobl, T. Blaschke & G. Griesebner (Eds.), Angewandte Geographische Informations-Verarbeitung XII (pp. 12-23). Wichmann Verlag, Karlsruhe, Germany.
[18] B. Waske, S. van der Linden, "Classifying Multilevel Imagery From SAR and Optical Sensors by Decision Funcion," IEEE Trans. Geosci. Remote Sens., vol. 46, no. 5, pp. 1457–1466, May 2008.