# SUPERPIXEL-BASED CLASSIFICATION OF HYPERSPECTRAL DATA USING SPARSE REPRESENTATION AND CONDITIONAL RANDOM FIELDS

*Ribana Roscher, Björn Waske*

Institute of Geographical Sciences, Freie Universität Berlin,
Malteserstr. 74–100, 12249 Berlin

**ABSTRACT**

This paper presents a superpixel-based classifier for landcover mapping of hyperspectral image data. The approach relies on the sparse representation of each pixel by a weighted linear combination of the training data. Spatial information is incorporated by using a coarse patch-based neighborhood around each pixel as well as data-adapted superpixels. The classification is done via a hierarchical conditional random field, which utilizes the sparse-representation output and models spatial and hierarchical structures in the hyperspectral image. The experiments show that the proposed approach results in superior accuracies in comparison to sparse-representation based classifiers that solely use a patch-based neighborhood.

***Index Terms***— Sparse coding, sparse representation, superpixel, hyperspectral, random field

## 1. INTRODUCTION

Hyperspectral imagery provides detailed and spectrally continuous spatial information on the land surface, while it is well known that increasing data dimensionality and high redundancy between features might cause problems during image analysis. Thus, different state-of-the-art classification methods have emerged over the past years, including support vector machines (SVM), ensemble based learning or classifiers based on multinomial logistic regression ([1], [2]). Recently, also sparse-representation based classifier, which use sparse coding strategies, have been introduced in the context of hyperspectral image classification. It could be shown that they achieve similar results as the state-of-the-art classifiers ([3], [4], [5]). Sparse coding strategies reconstruct a sample of interest as a sparse weighted linear combination of a few basis vectors chosen from a so-called dictionary. A dictionary is constructed either as the whole set of training vectors ([4], [6]) or by learning an adequate subset from the training vectors (e.g. [7], [8]).

One of the most important developments for sparse representation-based classifier for hyperspectral image classification is the incorporation of spatial information. This is either done via dictionary learning or via the sparse coding procedure, i.e. the estimation of the sparse parameter vector within the linear combination used for sparse representation. Many proposed approaches assume that remote sensing images are smooth, i.e. neighbored pixels tend to have similar spectral characteristic, and thus, exploit the spatial correlation within the sparse coding procedure ([5], [6]). However, such approaches only assumes a rectangular, i.e. patch-based, homogeneous neighborhood with similar spectral features. Therefore, actual class transitions, i.e. boundaries between classes within the image and regions of the same class, which show different spectral properties, cannot be considered. In order to mitigate the influence of this problem, e.g. [9] and [10] introduced different weights for all neighboring pixels depending on their similarity to the pixel of interest.

Instead of using a patch-based neighborhood, a more adapted neighborhood in terms of superpixels can be used. Superpixels, i.e. compact image segments composed of similar pixels, have recently emerged for remote sensing image analysis ([11], [12]) and show promising results regarding the classification accuracy. Since they group pixels of similar information they reduce the redundancy in the image and thus the size of the input for processing steps. While the patch-based joint sparsity model of [6] classifies each pixel separately, the superpixel-based approach classifies all pixels within a superpixel simultaneously. This makes the approach more efficient, but can also lead to misclassifications if the segmentation is not appropriate or the spectral features are not represented well by the training data. Moreover, since the sparse representation is performed only once per superpixel, generally, misclassifications affects a number of pixels depending on the superpixel size.

In this paper, the output of the superpixel-based sparse representation and the patch-based sparse representation are combined into a hierarchical conditional random field (CRF, [13, 14]). The combination proves to be beneficial because it considers information of a variety of neighbored pixels while at the same time being aware of object boundaries. The CRF enforces a smooth labeling by favoring neighboring pixel as well as neighboring superpixels to get the same class label. Moreover, a hierarchy is introduced in order connect the patch-based classification and the superpixel-based

The work was supported by German Research Foundation (DFG) WA 2728/3-1

classification and the pixels in it are pushed to get the same labels.

The following section describes the used methods comprising the sparse representation of images and the construction and optimization of the CRF. Sec. 3 comprises experiments to show performance of the proposed approach.

## 2. METHODS

The proposed classification procedure can be divided into 3 steps. In the first step a superpixel segmentation is obtained using the SLIC algorithm [15]. In the second step, the segmented image and the original image are both sparsely represented using the joint sparsity model of [6]. The class-wise reconstruction errors serve as input in the hierarchical CRF which provides the final labeling. The graph structure of the CRF is constructed by connecting neighbored pixels and superpixels and each superpixel with all pixels in it.

### 2.1. Sparse Coding

In terms of basic sparse coding a $(M \times 1)$-dimensional test sample $\boldsymbol{x}$ can be represented by a linear combination of a few training samples collected in a $(M \times D)$-dimensional dictionary $\boldsymbol{D}$, so that $\boldsymbol{x} = \boldsymbol{D}_1\boldsymbol{\alpha}_1 + \ldots + \boldsymbol{D}_K\boldsymbol{\alpha}_K = \boldsymbol{D}\boldsymbol{\alpha}$, whereas $\boldsymbol{\alpha}$ consists of the class-wise sparse parameter vectors ${}^P\boldsymbol{\alpha}_k$ with $k \in \{1, \ldots, K\}$, which are multiplied with the class-wise sub-dictionaries $\boldsymbol{D}_k$. The optimization problem is given by

$$\hat{\boldsymbol{\alpha}} = \operatorname{argmin} \|\boldsymbol{D}\boldsymbol{\alpha} - \boldsymbol{x}\|_b \quad \text{subject to} \quad \|\hat{\boldsymbol{\alpha}}\|_0 < W, \tag{1}$$

yielding a sparse parameter vector $\hat{\boldsymbol{\alpha}}$, whereas $W$ is the number of nonzero elements and $b$ specifies the used norm.

In order to incorporate neighbored pixels the reconstruction step is extended to a joint sparsity model so that the pixel of interest and their neighbors $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]$ are reconstructed with the same dictionary elements, however, allowing for different weights $\boldsymbol{A} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N]$. The set of all pixels within the patch or the superpixel is given by $\mathcal{N}$ and the number of elements is $N$. The optimization problem is than given by

$$\hat{\boldsymbol{A}} = \operatorname{argmin} \|\boldsymbol{D}\boldsymbol{A} - \boldsymbol{X}\|_F \quad \text{subject to} \quad \|\hat{\boldsymbol{A}}\|_{\text{row},0} < K, \tag{2}$$

where $\|\hat{\boldsymbol{A}}\|_{\text{row},0}$ is the number of nonzero rows and $F$ is the Frobenius norm. E.g. the problem can be solved with SOMP [6, 16].

The class-wise reconstruction error of the test sample $\boldsymbol{x}$ is derived by

$$r(k, \boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{D}^k\hat{\boldsymbol{\alpha}}^k\|_2. \tag{3}$$

Ideally the test sample $\boldsymbol{x}$ belonging to class $k$ can be reconstructed best by using only dictionary elements from $\boldsymbol{D}^k$. The mean reconstruction error

$$\overline{r}(k, \boldsymbol{X}) = \frac{1}{N}\sum_{i\in\mathcal{N}} \|\boldsymbol{x}_i - \boldsymbol{D}^k\hat{\boldsymbol{\alpha}}_i^k\|_2. \tag{4}$$

Both serve as input in the CRF.

### 2.2. Conditional Random Field

Given an image $\boldsymbol{I}$ that contains $J$ pixel and $S$ superpixel, the task is to classify each superpixel $\boldsymbol{X}_s$ with $s \in \{1, \ldots, S\}$. A hierarchical CRF is employed in order to model prior knowledge about the neighborhood relations of the pixels and superpixels within the image and between the layers, whereas one layer contains the pixel and one layer contains the superpixel. Within the CRF graph a superpixel is connected to all its neighors and to all pixels which lie in the superpixel. Additionally, each pixel is connected to its neighbors. The best patch-based classification ${}^Py$ and superpixel-based classification ${}^Sy$ based on the image data $\boldsymbol{X}$ is estimated by the argument of the minimum of the energy

$$\begin{aligned} E({}^Py, {}^Sy) &= \sum_j r({}^Py_j, \boldsymbol{x}_j) + \delta \sum_s \overline{r}({}^Sy_s, \boldsymbol{X}_s) \\ &- \beta \sum_{\{j,l\}\in {}^PQ} \cos(\boldsymbol{x}_j, \boldsymbol{x}_l)\delta\left({}^Py_j, {}^Py_l\right) \\ &- \beta \sum_{\{s,t\}\in {}^SQ} \cos(\overline{\boldsymbol{x}}_s, \overline{\boldsymbol{x}}_t)\delta\left({}^Sy_s, {}^Sy_t\right) \\ &- \gamma \sum_{\{s,j\}\in {}^LQ} \cos(\overline{\boldsymbol{x}}_s, \boldsymbol{x}_j)\delta\left({}^Sy_s, {}^Py_j\right), \end{aligned} \tag{5}$$

where the first two terms are the unary terms and the rest are the binary terms, $\delta$ is the Kronecker delta function, $\overline{\boldsymbol{x}}_s$ is the mean feature of all pixels in the $s$-th superpixel, ${}^SQ$ is the set of all neighboring superpixels, ${}^PQ$ is the set of all neighboring pixels and ${}^LQ$ is the set of the neighboring pixels and superpixels between the layers, i.e. the incidence relation between pixels and superpixels. The weighting parameters are given by $\delta$, $\beta$ and $\gamma$, which can be determined via cross-validation.

## 3. EXPERIMENTAL SETUP AND RESULTS

### 3.1. Data sets

The considered datasets are UNIVERSITY OF PAVIA and INDIAN PINES. The UNIVERSITY OF PAVIA dataset was acquired by ROSIS and covers $610 \times 340$ pixels, with a spatial resolution of 1.3 m. Some bands have been removed due to noise, the remaining 103 bands have been used in the classification. The classification is aiming on nine land cover classes. The INDIAN PINES dataset was acquired by AVIRIS and covers $145 \times 145$ pixels, with a spatial resolution of 20 m and 224 bands. The reference data consist 16 classes. The training data is randomly selected and comprises about $10\%$ of the labeled data (see Tab. 1). Each channel of the images were normalized to zero mean and standard deviation of 1. Moreover, each pixel were normalized to have a unit length of 1.

### 3.2. Experimental setup

Each image is sparsely represented using the methods presented in Sec. 2.1 and classified via a hierarchical conditional random field (see Sec. 2.2). The superpixels are derived using the SLIC approach of [15]. Different size of the patch-based neighboorhood and superpixels are analysed by means of the classification accuracy. The optimization of the energy in the CRF is done via graph-cut [17]. The CRF parameters $\beta$ and $\gamma$ are empirically determined via crossvalidation, whereas the parameter $\delta$ is set to zero showing better results than using both unary terms. In all experiments the maximum number of used dictionary elements is fixed to $W = 5$. The experiments compare the support vector machines with composite kernel (SVMCK, [18]), SOMP with patch-based neighborhood [6] and the combined patch-based and superpixel-based SOMP with hierarchical CRF (CSOMP-CRF).

### 3.3. Results and discussion

Table 1 shows the average results of all runs with the best parameter settings. It indicates that the superpixel-based approaches achieve higher average accuracies than the patch-based SOMP, kernel-based SOMP and SVMCK. The combined approach with CRF achieves also a better overall accuracy and Kappa coefficient than the patch-based SOMP and SVMCK and competitive results to the kernel SOMP. Fig. 1 demonstrates the influence of the patchsize $Z_{\text{p}} \times Z_{\text{p}}$ and the roughly size of the superpixels $Z_{\text{sp}} \times Z_{\text{sp}}$. It can be seen that small superpixel sizes of $Z_{sp} = 3$ to $Z_{sp} = 7$ achieve the best results. Small patchsizes for the SOMP provide noisy results and too large patchsizes provide indistinct classification results, both indicated by low accuracies. Thus, the patchsize and the superpixel size have to be chosen adequately depending on the image structure.

The results for the University of Pavia dataset show similar findings as for the Indian Pines dataset. Using the best parameter setting an overall accuracy of $87.04\%$, an average accuracy of $89.09\%$ and a Kappa coefficient of $0.897$ could be obtained. This is significantly better than for SOMP, showing an overall accuracy of $79.00\%$, an average accuracy of $86.04\%$ and a Kappa coefficient of $0.728$, and kernel-based SOMP, showing an overall accuracy of $85.67\%$, an average accuracy of $85.83\%$ and a Kappa coefficient of $0.815$.

**Table 1**. Size of training and test data, classwise accuracies, overall accuracy (oa), average accuracy (aa) and Kappa coefficient ($\kappa$) of Indian Pines dataset using support vector machines with composite kernel (SVMCK), simultaneous orthogonal matching pursuit (SOMP), kernel simultaneous orthogonal matching pursuit (KSOMP) and the combined patch-based and superpixel-based SOMP with hierarchical CRF (CSOMP-CRF).

| | # train | # test | SVMCK | SOMP | KSOMP | CSOMP-CRF |
|---|---|---|---|---|---|---|
| Alfalfa | 6 | 48 | 95.83 | 85.42 | **97.92** | 88.10 |
| Corn-notill | 144 | 1290 | 96.67 | 94.88 | **97.21** | 94.74 |
| Corn-min | 84 | 750 | 90.93 | 94.93 | **96.67** | 94.63 |
| Corn | 24 | 210 | 85.71 | 91.43 | 93.33 | **95.92** |
| Grass/Pasture | 50 | 447 | 93.74 | 89.49 | **95.75** | 94.60 |
| Grass/Trees | 75 | 672 | 97.32 | 98.51 | **99.55** | 97.51 |
| Grass-mowed | 3 | 23 | 69.57 | 91.30 | 60.78 | **98.76** |
| Hay | 49 | 440 | 98.41 | 99.55 | **100.00** | 99.45 |
| Oats | 2 | 18 | **55.56** | 0.00 | 0.00 | 47.62 |
| Soybeans-notill | 97 | 871 | 93.80 | 89.44 | 94.60 | **95.51** |
| Soybeans-min | 247 | 2221 | 94.37 | 97.34 | **99.28** | 98.63 |
| Soybeans-clean | 62 | 552 | 93.66 | 88.22 | **95.65** | **95.65** |
| Wheat | 22 | 190 | 99.47 | **100.00** | **100.00** | 99.62 |
| Woods | 130 | 1164 | 99.14 | 99.14 | 99.83 | **100.00** |
| Building-Grass | 38 | 342 | 87.43 | **99.12** | 91.81 | 98.83 |
| Stone | 10 | 85 | **100.00** | 96.47 | 91.76 | 99.16 |
| Overall | 1043 | 9323 | 94.86 | 95.28 | **97.33** | 97.06 |
| Average | | | 90.73 | 88.45 | 88.39 | **93.67** |
| $\kappa$ | | | 0.941 | 0.946 | **0.970** | 0.966 |

## 4. CONCLUSION

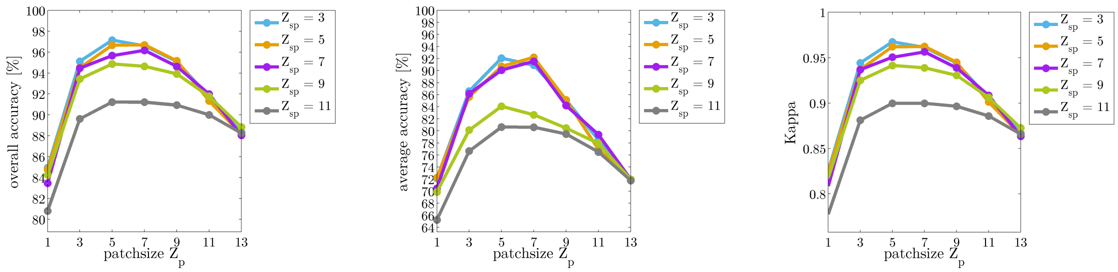
The paper presented a sparse-representation based classifier which combines a patch-based classification and a superpixel-based classification in a hierarchical conditional random field. The experimental results underline that the approach achieve better results than sparse-representation based classifier which solely use patch-based spatial information.

## 5. REFERENCES


[1] Gustavo Camps-Valls, Devis Tuia, Lorenzo Bruzzone, and J Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2014.

[2] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, 2013.

[3] Benqin Song, Jun Li, M. Dalla Mura, Peijun Li, A. Plaza, J.M. Bioucas-Dias, J. Atli Benediktsson, and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute




(a) Overall accuracy (b) Average accuracy (c) Kappa coefficient

**Fig. 1**. Influence of the patchsize $Z_p$ and the superpixel size $Z_{sp}$ onto the overall accuracy, average accuracy and Kappa coefficient for the INDIAN PINES dataset.

profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, 2014.

[4] Ali-Asghar Soltani-Farani, Hamid R. Rabiee, and Seyyed Abbas Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *Proc. Conference on Computer Vision and Pattern Recognition*, 2013.

[5] Y. Chen, Nasser M. Nasrabadi, and Trac D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, 2013.

[6] Y. Chen, Nasser M. Nasrabadi, and Trac D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, 2011.

[7] Shuyuan Yang, Honghong Jin, Min Wang, Yu Ren, and Licheng Jiao, "Data-driven compressive sampling and learning sparse coding for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 479–483, 2014.

[8] Adam S Charles, Bruno A Olshausen, and Christopher J Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 963–978, 2011.

[9] Haoliang Yuan, Yang Lu, Lina Yang, Huiwu Luo, and Yuan Yan Tang, "Sparse representation using contextual information for hyperspectral image classification," in *IEEE International Conference on Cybernetics*, 2013, pp. 138–143.

[10] U. Srinivas, Yi Chen, V. Monga, N.M. Nasrabadi, and T.D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, 2013.

[11] Jingting Liu, Weidong Yang, Shulun Tan, and Zijian Wang, "Remote sensing image classification based on random projection super-pixel segmentation," in *Proc. Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, 2013, vol. 8921.

[12] J. Stefanski, B. Mack, and B. Waske, "Optimization of object-based image analysis with random forests for land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 6, no. 6, pp. 2492–2504, 2013.

[13] S. Kumar and M. Hebert, "Discriminative random fields," *International Journal of Computer Vision*, vol. 68, no. 2, pp. 179–201, 2006.

[14] John D. Lafferty, Andrew McCallum, and Fernando C. N. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proc. International Conference on Machine Learning*, 2001, pp. 282–289.

[15] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "Slic superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, 2012.

[16] Joel A Tropp, Anna C Gilbert, and Martin J Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[17] Yuri Boykov, Olga Veksler, and Ramin Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, 2001.

[18] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muñoz-Marí, Joan Vila-Francés, and Javier Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, 2006.