Shapelet-Based Sparse Image Representation for Landcover Classification of Hyperspectral Data

Ribana Roscher, Björn Waske

Department of Earth Sciences, Institute of Geographical Sciences

Freie Universität Berlin

Abstract—This paper presents a novel sparse representation-based classifier for landcover mapping of hyperspectral image data. Each image patch is factorized into segmentation patterns, also called shapelets, and patch-specific spectral features. The combination of both is represented in a patch-specific spatial-spectral dictionary, which is used for a sparse coding procedure for the reconstruction and classification of image patches. Hereby, each image patch is sparsely represented by a linear combination of elements out of the dictionary. The set of shapelets is specifically learned for each image in an unsupervised way in order to capture the image structure. The spectral features are assumed to be the training data. The experiments show that the proposed approach shows superior results in comparison to sparse-representation based classifiers that use no or only limited spatial information and behaves competitive or better than state-of-the-art classifiers utilizing spatial information and kernelized sparse representation-based classifiers.

I. INTRODUCTION

Hyperspectral image classification has become a valuable instrument for environmental monitoring and the analysis of geoscientific data ([1], [2]). Especially supervised classification methods such as support vector machines (SVM, [3], [4]), ensemble based learning ([5], [6]) or classifiers based on multinomial logistic regression ([7], [8]) have been proven to be powerful tools for classification.

A successful development regarding these classifiers is the additional integration of spatial/contextual information in order to take advantage of the correlation between spatially adjacent pixels (c.f. [1]). E.g. [9] and [10] introduce SVM with additional spatial information by means of using a composite kernel. An alternative approach was proposed by e.g. [11] and [12], which is based on mathematical morphology. Both approaches were combined by [13] to generalized composite kernels for multinomial logistic regression. Other spectral-spatial approaches are based on markov random fields (e.g. [14], [15]) or the classification of pre-defined regions, which are obtained by image segmentation (e.g. [16], [17]).

Besides the aforementioned classifiers, sparse representation-based classifiers have been recently introduced in the context of hyperspectral image classification (e.g. [18], [19], [20], [21]). They must not to be confused with sparse classifiers such as SVM. Sparse representation-based classifier have shown a more accurate or at least equally well performance than powerful state-of-the-art classifier, among others because they are also able to use spectral as well as spatial information. A sparse representation assumes that each pixel can be reconstructed by a sparsely weighted linear combination of a few basis vectors, taken from a so-called dictionary. The dictionary is constructed from a set of representative samples, such as the training data, and is either directly embodied by these samples (e.g. [20], [21]) or learned from them (e.g. [22], [23], [24]). In the context of supervised classification each dictionary element also provides a class label, which is used for the classification of the reconstructed sample of interest. Formally, the dictionary is chosen to be overcomplete, i.e. the number of dictionary elements is larger than the dimension of elements. For the reconstruction and classification of RGB- and gray-valued image data, generally, the dictionary elements are representative vectorized image patches, which are derived from labeled training data (e.g. [25], [26]). In this way, spectral as well as spatial information are integrated in the dictionary. However, this approach is inapplicable for hyperspectral data, due to the number and dimension of dictionary elements as well as the usually relatively limited amount of training data.

Recently, this problem has been considered and various sparse-representation based classifiers have been proposed which incorporate spectral and spatial information in an alternative way. 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. Various 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. E.g. [27] successfully applied a joint sparsity model for hyperspectral image classification, where neighbored pixels within an image patch are sparsely represented by a common set of dictionary elements, however, allowing for different weights in the sparse linear combination. The model is realized via simultaneous orthogonal matching pursuit (SOMP, [28]), an extension of the orthogonal matching pursuit (OMP, [29]), which is usually used for solving the sparse coding task. In these approaches, the dictionary is assumed to be the training data and thus, the dictionary elements are treated independently. In contrast to this, [20] learn the dictionary elements using the joint sparsity assumption like [27] and so-called contextual groups, i.e. non-overlapping image patches. The estimated sparse parameter vectors are considered as features for a linear SVM classifier. However, the mentioned approaches only assumes homogeneous regions with similar spectral features in each patch. 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. [19] and [30] introduced different weights for all neighboring pixels depending on their similarity to the pixel of interest.

This paper presents a novel sparse-representation based classifier that explicitly introduces prior knowledge about the spatial nature of the image. Therefore, it is more comprehensive than approaches that only assume a homogeneous neighborhood. The approach is used for hyperspectral image data and based on the work of [31], who applied a similar procedure for the recognition of objects in RGB-images. Each image patch is factorized into segmentation patterns, also called shapelets, and patch-specific spectral feature palettes which describe how to color the shapelets. The shapelets, organized in a shapelet-dictionary, are a collection of patches containing grouping of pixels that tend to co-occur in its spectral features. The elements of a representative spectral feature palette, in the simplest case the training data, are organized in a spectral dictionary. The shapelets and a well-chosen patch-specific spectral palette, a subset of the whole spectral palette, are combined to a patch-specific spatial-spectral dictionary, which is used for a sparse coding procedure for the reconstruction of the image patches. Hereby, it is assumed that each patch can be represented by a sparsely weighted linear combination of elements out of this patch-specific dictionary. The estimated sparse weights and their assigned dictionary elements are then used for classification of the patch.

The following section describes the shapelet-based sparse representation scheme and the used methods comprising the learning of the shapelets, the derivation of the patch-specific spatial-spectral dictionary and the sparse coding procedure. Sec. III comprises experiments to show performance of the proposed approach.

II. METHODS

Given an image I that contains J overlapping patches of size $Z \times Z$, the task is to classify each patch X_j with $j \in \{1, \ldots, J\}$ and combine the results to a classified image. Each patch is represented via a sparsely weighted linear combination, whereas the sparse weighting vector is determined using a sparse coding procedure (see Sec. II-B). The procedure utilizes a set of learned shapelets $\mathcal{S}$ (see Sec. II-C) and a spectral set $\mathcal{F}$, which is called spectral palette according to [31] (see Sec. II-D).

A. Shapelet-Based Classification Scheme

Fig. 1 illustrates the shapelet-based classification scheme used in this paper. Each image patch can be represented by a selected set of shapelets and spectral features from a patch-specific palette. The set of shapelets, organized in a shapelet dictionary, is learned from each image and the spectral palette is assumed to be the training data. A patch-specific spectral palette (a subset of the whole spectral palette) is derived for each patch and used for the coloring of the shapelets yielding a patch-specific spatial-spectral dictionary. Each pixel in the patch is finally classified by a voting scheme explained in Sec. II-F.

B. Basic Sparse Coding

In terms of basic sparse coding a $(M \times 1)$-dimensional test sample x can be represented by a linear combination of a few training samples collected in a $(M \times D)$-dimensional dictionary D, so that $x = D^1\alpha^1 + \ldots + D^K\alpha^K = D\alpha$, whereas
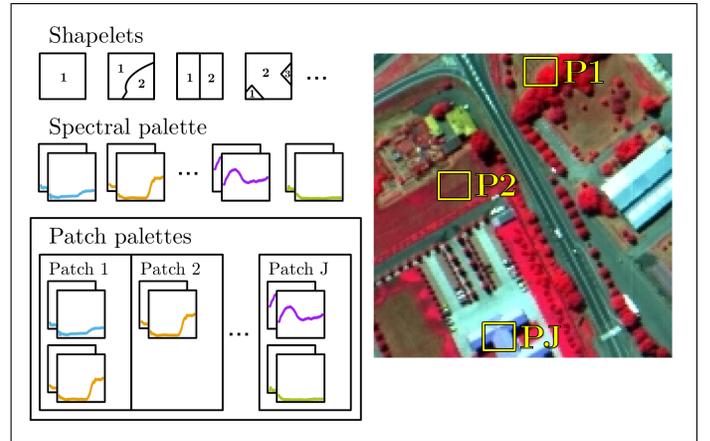

Fig. 1: Shapelet-based classification scheme illustrated by means of a hyperspectral image. Each image patch can be described by a set of learned shapelets with varying number of regions (denoted with numbers), which are colored by a patch-specific palette comprising specific spectral features. Each patch palette is a subset of the whole spectral palette, which is assumed to be the labeled training data (different colors of the signatures are given by different classes).

α consists of the class-wise sparse parameter vectors α^k with $k \in \{1, \ldots, k, \ldots, K\}$, which are multiplied with the class-wise sub-dictionaries D^k. The optimization problem is given by

$$\hat{\alpha} = \arg\min \|D\alpha - x\|_b \quad \text{subject to} \quad \|\hat{\alpha}\|_0 < W, \quad (1)$$

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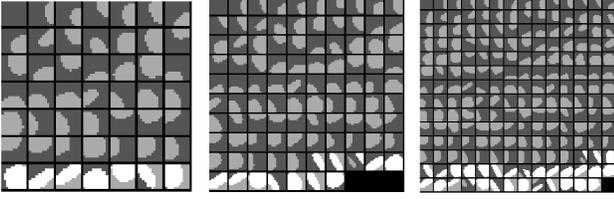
C. Shapelet Dictionary

This step describes the extraction of the most representative segmentation patterns (shapelets). The $Z \times Z$ shapelets $S_p \in \mathcal{S}$ with $p \in \{1, \ldots, P\}$, each containing R_p regions, are specifically learned for each image. The shapelets are vectorized and collected in a $((Z \cdot Z) \times P)$-dimensional matrix $^S D$ called shapelet dictionary. In order to extract image-specific segmentation patterns, firstly the SLIC superpixel segmentation approach of [32] is applied to the image. It provides a segmentation of the image with compact segments containing homogeneous spectral features. It is assumed that borders between classes in the image coincide with the superpixel borders. Given the segmented image, patches of size $Z \times Z$ are extracted and converted to binary patches. In detail, for each region a binary patch is created with the region's pixels set to 1 and the remaining pixel to 0.

The most representative segmentation patterns out of the extracted patches are found by an integer k-means clustering [33], whereas the means are defined to be the elements of the shapelet dictionary. Fig. 2 shows extracted shapelet dictonaries of different size with at most 3 regions.

D. Spectral Dictionary

This step introduces the spectral palette, derived from the training data. The training set consists of L labeled sam-

(a) Small shapelet dictionary from segments of a roughly size of 10×10

(b) Medium shapelet dictionary from segments of a roughly size of 10×10

(c) Large shapelet dictionary from segments of a roughly size of 10×10

Fig. 2: Visualized shapelet dictionaries of different sizes consisting of shapelets with at most 3 regions. The small dictionary results from 50 extracted means (from clustering), the medium sized dictionary from 100 means and the large dictionary from 200 means. Some means were removed if they had more than 3 regions or if they are redundant.

ples $(\boldsymbol{x}_l, y_l) \in \mathcal{L}$ with $l \in \{1, \ldots, L\}$ with m-dimensional spectral feature vectors $\boldsymbol{x}_l \in \mathrm{IR}^M$ and class labels $y_l \in \{1, \ldots, k, \ldots, K\}$. The N_k training samples of class k are denoted with $\boldsymbol{x}_{k,l}$. All training samples are collected in a $(M \times L)$-dimensional structured matrix ${}^L D = \begin{bmatrix} {}^L D_1 \ldots {}^L D_K \end{bmatrix} = [[\boldsymbol{x}_{1,1} \ldots \boldsymbol{x}_{1,N_1}] \ldots [\boldsymbol{x}_{K,1} \ldots \boldsymbol{x}_{K,N_K}]]$, called spectral dictionary, whereas ${}^L D_k$ denotes the class-specific spectral sub-dictionary.

E. Patch-Specific Spatial-Spectral Dictionary

In order to solve the classification task via sparse coding while using a dictionary with both spatial and spectral information, the shapelet dictionary and the spectral dictionary are combined to a patch-specific spatial-spectral dictionary ${}^C D_j \left({}^S D, {}^L D, \boldsymbol{X}_j \right)$ for the j-th patch rather than learning a fixed dictionary for the whole image. The basic idea for the determination of the spatial-spectral dictionary element given a specific shapelet is to minimize the distance to a given image patch by using an arbitrary distance measure. This is done by deriving a patch-specific spectral palette and "filling" the shapelet with these elements. In this paper, the patch-specific spectral palette contains all spectral features from at most R_p classes (i.e. the number of regions in the p-th shapelet), which represent the regions best. The most representing class for a region can be found in several ways. For example, if the training samples are spatial evenly distributed over the image, it has shown to be powerful when the most representing class of the region is assumed to have the label of the sample with minimum distance. Ideally, this sample is a training sample. Another possibility is to first classify all pixels in this region individually, e.g. by nearest neighbor, and then choosing the dominant class by majority vote. After finding the best fitting elements from the patch-specific spectral palette for each pixel in the image patch, the spatial-spectral dictionary element is vectorized, i.e. it has a dimension of $(Z \cdot Z \cdot M) \times 1$, and collected in the dictionary. Unlike the structured spectral dictionary introduced in Sec. II-D each element in ${}^C D_j$ can belong to different classes, which must be taken into account for the classification step.

The sparse coding procedure for the reconstruction of a patch $\boldsymbol{X}_j$ is than given by

$$\hat{\boldsymbol{\alpha}}_j = \operatorname{argmin} \|{}^C D_j \boldsymbol{\alpha}_j - \boldsymbol{x}_j\|_b \qquad \text{subject to} \qquad \|\hat{\boldsymbol{\alpha}}_j\|_0 < W, \tag{2}$$

whereas $\boldsymbol{x}_j = \operatorname{vec}(\boldsymbol{X}_j)$ is the vectorized image patch. Please note that this dictionary generally is not overcomplete anymore.

F. Classification

Once the parameters for the sparse representation of the patch $\boldsymbol{x}_j$ are estimated, class-votes for each pixel in the patch can be derived by the following procedure. Since the patches are overlapping, one pixel provides several votes contributing to a final classification of the whole image.

The votes for class k of the t-th pixel $x_{j,t}$ with $t \in \{1, \ldots, (Z \cdot Z)\}$ in the j-th patch can be obtained by $v_{j,t}^k = 1/r_{j,t}^k$, where $r_{j,t}^k$ is the reconstruction error given by

$$r_{j,t}^k = \|x_{j,t} - {}^C D_{j,t}^k \hat{\boldsymbol{\alpha}}_{j,t}^k\|_2, \tag{3}$$

whereas ${}^C D_{j,t}^k$ are all dictionary entries which belong to class k. The parameters $\hat{\boldsymbol{\alpha}}_{j,t}^k$ are the weights assigned to the dictionary elements ${}^C D_{j,t}^k$.

III. EXPERIMENTS

A. Data Sets

The considered data sets are two hyperspectral images - UNIVERSITY OF PAVIA and INDIAN PINES - from study sites with different environmental setting. The INDIAN PINES dataset was acquired by AVIRIS and covers 145×145 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 Fig. 4 and Tab. I). The UNIVERSITY OF PAVIA dataset was acquired by ROSIS and covers 610×340 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 nine land cover classes (see Fig. 4 and Tab. II). Each channel of the images was normalized to zero mean and standard deviation of 1. Moreover, each pixel were normalized to have a unit length of 1.

B. Experimental Setup

Each image is sparsely represented using the methods presented in Sec. II-E and classified using (3). The spatial-spectral dictionary is constructed using the inverse correlation coefficients as distances measure showing better results than commonly used distance measures such as L_1-, L_2- or $L_{\inf}$-norm. The number of shapelets and the patch size is varied for the INDIAN PINES dataset and their influence onto the classification result is analyzed. In all experiments the maximum number of used dictionary elements is fixed to $W = 3$. The results of the presented approach is compared to support vector machines with composite kernel (SVMCK, [9]), the simultaneous orthogonal matching pursuit (SOMP, [27]), simultaneous subspace pursuit (SSP, [27]) and the best kernelized sparse coding algorithm presented in [21]. The results of SVMCK, SOMP and SSP were also taken from [21]. Moreover, the presented approach is compared to the sparse

coding approach with spectral-contextual dictionary learning (SCDL) presented by [20].

C. Results and Discussion

1) INDIAN PINES *Dataset:* The accuracy measures for ten different results, achieved by the best parameter setting (see Fig. 3), are shown in Tab. I. The training data, test data and classification map of the run with the highest average accuracy using the best parameter setting is shown in Fig. 4. The results show that the proposed approach achieves better results than SVMCK, SOMP and SSP and comparable results to SCDL and kernel subspace pursuit with composite kernel (KSPCK).

TABLE I: Size of training and test data, classwise accuracies, overall accuracy (oa), average accuracy (aa) and kappa coefficient (κ) of INDIAN PINES dataset using support vector machines with composite kernel (SVMCK), simultaneous orthogonal matching pursuit (SOMP), simultaneous subspace pursuit (SSP), kernel subspace pursuit with composite kernel (KSPCK), sparse coding approach with spectral-contextual dictionary learning (SCDL) and the shapelet-based sparse coding approach (SSC).

	# train	# test	SVMCK	SOMP	SSP	KSPCK	SCDL	SSC
Alfalfa	6	48	95.83	85.42	81.25	95.83	93.75	**96.53**
Corn-notill	144	1290	96.67	94.88	95.74	**99.15**	94.93	96.73
Corn-min	84	750	90.93	94.93	92.80	96.93	97.39	**98.11**
Corn	24	210	85.71	91.43	82.38	97.14	90.57	**98.10**
Grass/Pasture	50	447	93.74	89.49	93.29	98.21	97.23	**98.32**
Grass/Trees	75	672	97.32	98.51	98.81	99.11	99.17	**99.80**
Grass/Pasture-mowed	3	23	69.57	91.30	82.61	**100.00**	**100.00**	94.93
Hay-windrowed	49	440	98.41	99.55	99.77	**99.97**	99.95	99.81
Oats	2	18	55.56	0.00	0.00	**100.00**	79.44	95.37
Soybeans-notill	97	871	93.80	89.44	91.27	**97.70**	96.30	97.57
Soybeans-min	247	2221	94.37	97.34	97.43	98.20	98.46	**99.53**
Soybeans-clean	62	552	93.66	88.22	89.13	**98.73**	92.97	96.32
Wheat	22	190	99.47	**100.00**	99.47	**100.00**	99.05	99.91
Woods	130	1164	99.14	99.14	99.05	99.48	98.87	**99.90**
Building-Grass-Trees	38	342	87.43	**99.12**	97.95	97.37	97.13	98.20
Stone-steel Towers	10	85	**100.00**	96.47	92.94	95.29	96.00	95.69
Overall	1043	9323	94.86	95.28	95.34	98.47	97.81	**98.53**
Average			90.73	88.45	87.12	**98.31**	95.70	97.80
κ			0.941	0.946	0.947	**0.983**	0.968	**0.983**

Fig. 3 demonstrates the influence of the number of shapelets and the patch size on the classification accuracies. The results show that the highest accuracies are obtained with a patch size of 9×9 pixel. The best overall accuracy is obtained with 10 shapelets, the highest average accuracy with 25 shapelets and the best kappa coefficient with 10 shapelets. The plots clearly indicate the gain of using shapelets rather than using only one homogeneous patch, i.e. one shapelet. Especially for large patch sizes the accuracies significantly increase with an increasing number of shapelets. However, if the patch size is small an increased number of shapelets results in lower accuracies. This is caused by an overfitting effect, whereas noise is fitted by potentially non-representative shapelets. The plots also show that even with only one homogeneous patch the proposed framework can achieve higher accuracies than many of the considered approaches. This underlines the fact that it is worth to learn a specifically designed dictionary in order to increase the classification accuracy.

2) UNIVERSITY OF PAVIA *Dataset:* The accuracy measures are presented in Table II and the landcover map is presented in Fig. 4. The results clearly show that the incorporation of shapelets significantly improves the overall classification accuracy and the kappa coefficient. The average accuracy is competitive to the best kernel based approach, namely kernel simultaneous subspace pursuit (KSSP), and significantly higher than the other considered approaches. The results could be obtained with a relatively large patch size of 13×13 pixel and thus, the classification result is very smooth within large homogeneous regions such as MEADOWS. Nevertheless, also small regions such as SHADOWS and METAL SHEETS can be classified well, which is mainly due to the usage of shapelets. Besides the proposed approach also SCDL uses a large patch size of 16×16 pixel, while all other approaches (SVMCK, SOMP, SSP and KSSP) use a smaller patch size of 5×5 pixel ([20], [21]).

TABLE II: Size of training and test data, classwise accuracies, overall accuracy (oa), average accuracy (aa) and kappa coefficient (κ) of UNIVERSITY OF PAVIA dataset using support vector machines with composite kernel (SVMCK), simultaneous orthogonal matching pursuit (SOMP), simultaneous subspace pursuit (SSP), kernel subspace pursuit with composite kernel (KSPCK), sparse coding approach with spectral-contextual dictionary learning (SCDL) and the shapelet-based sparse coding approach (SSC).

	# train	# test	SVMCK	SOMP	SSP	KSSP	SCDL	SSC
Asphalt	548	6304	79.85	59.33	69.59	89.56	81.87	**92.66**
Meadows	540	18146	84.86	78.15	72.31	79.98	96.48	**98.46**
Gravel	392	1815	81.87	83.53	74.10	**85.45**	83.36	73.88
Trees	524	2912	96.36	96.91	95.33	**98.66**	95.47	95.47
Metal sheets	265	1113	99.37	99.46	99.73	99.91	99.82	**100.00**
Bare soil	532	4572	93.55	77.41	86.72	**95.76**	81.21	83.81
Bitumen	375	981	90.21	98.57	90.32	97.96	74.11	**99.90**
Bricks	514	3364	92.81	89.09	90.46	**96.43**	85.91	95.99
Shadows	231	795	95.35	91.95	90.94	98.49	96.60	**99.37**
Overall	3921	40002	87.18	79.00	78.39	87.65	90.42	**94.36**
Average			90.47	86.04	85.50	**93.58**	88.31	93.21
κ			0.833	0.728	0.724	0.840	0.870	**0.923**

IV. CONCLUSION

The paper presents a shapelet-based sparse representation-based approach for the classification of hyperspectral image data. The experimental results underline that the approach is competitive to state-of-the-art classifiers, which use spatial information, and is superior to classification procedures which use no or only limited spatial information. Future work could address a more sophisticated learning of the shapelet dictionary and spectral dictionary, such as discriminative dictionaries. In addition the classification step can be replaced by using a classifier such as support vector machines with the sparse parameters vectors as features.
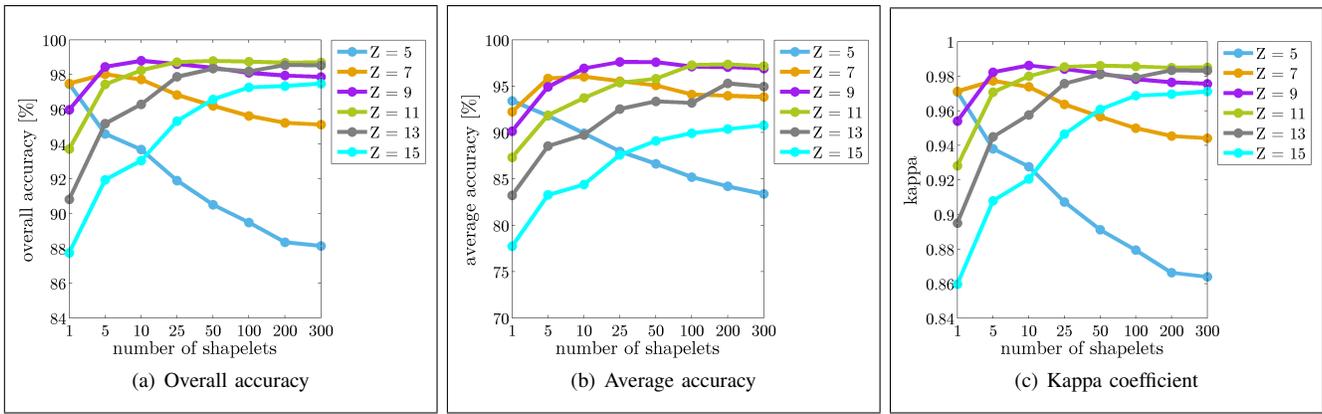
Fig. 3: Influence of the number of shapelets and the patch size onto the overall accuracy, average accuracy and kappa coefficient for the INDIAN PINES dataset.

REFERENCES

[1] G. Camps-Valls, D. Tuia, L. 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 Geoscience and Remote Sensing Magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[3] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[4] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, 2004.

[5] B. Waske, S. van der Linden, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2880–2889, 2010.

[6] J. C.-W. Chan and D. Paelinckx, "Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2999–3011, 2008.

[7] R. Roscher, B. Waske, and W. Förstner, "Incremental import vector machines for classifying hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 9, no. 99, pp. 3463–3473, 2012.

[8] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.

[9] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 3, no. 1, pp. 93–97, 2006.

[10] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multisource composite kernels for urban-image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 1, pp. 88–92, 2010.

[11] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 11, pp. 3804–3814, 2008.

[12] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 10, pp. 3747–3762, 2010.

[13] J. Li, P. Reddy Marpu, A. Plaza, J. Bioucas-Dias, and J. Atli Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4816–4829, Sept 2013.

[14] W. Li, S. Prasad, and J. Fowler, "Hyperspectral image classification using gaussian mixture models and markov random fields," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 153–157, Jan 2014.

[15] G. Moser and S. B. Serpico, "Combining support vector machines and markov random fields in an integrated framework for contextual image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 5, pp. 2734–2752, 2013.

[16] J. Stefanski, B. Mack, and B. Waske, "Optimization of object-based image analysis with random forests for land cover mapping," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 6, pp. 2492–2504, Dec 2013.

[17] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 1, pp. 2–16, 2010.

[18] B. Song, J. Li, M. Dalla Mura, P. Li, A. Plaza, J. Bioucas-Dias, J. Atli Benediktsson, and J. Chanussot, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 5122–5136, Aug 2014.

[19] H. Yuan, Y. Lu, L. Yang, H. Luo, and Y. Y. Tang, "Sparse representation using contextual information for hyperspectral image classification," in *IEEE International Conference on Cybernetics*. IEEE, 2013, pp. 138–143.

[20] A.-A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[21] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 217–231, Jan 2013.

[22] S. Yang, H. Jin, M. Wang, Y. Ren, and L. Jiao, "Data-driven compressive sampling and learning sparse coding for hyperspectral image classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 479–483, 2014.

[23] A. S. Charles, B. A. Olshausen, and C. J. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 963–978, 2011.

[24] A. Castrodad, Z. Xing, J. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4263–4281, 2011.

[25] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 4, pp. 791–804, April 2012.

[26] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods

(a) Image   (b) Training data   (c) Test data   (d) Result with SSC

(e) Image   (f) Training data   (g) Test data   (h) Result with SSC
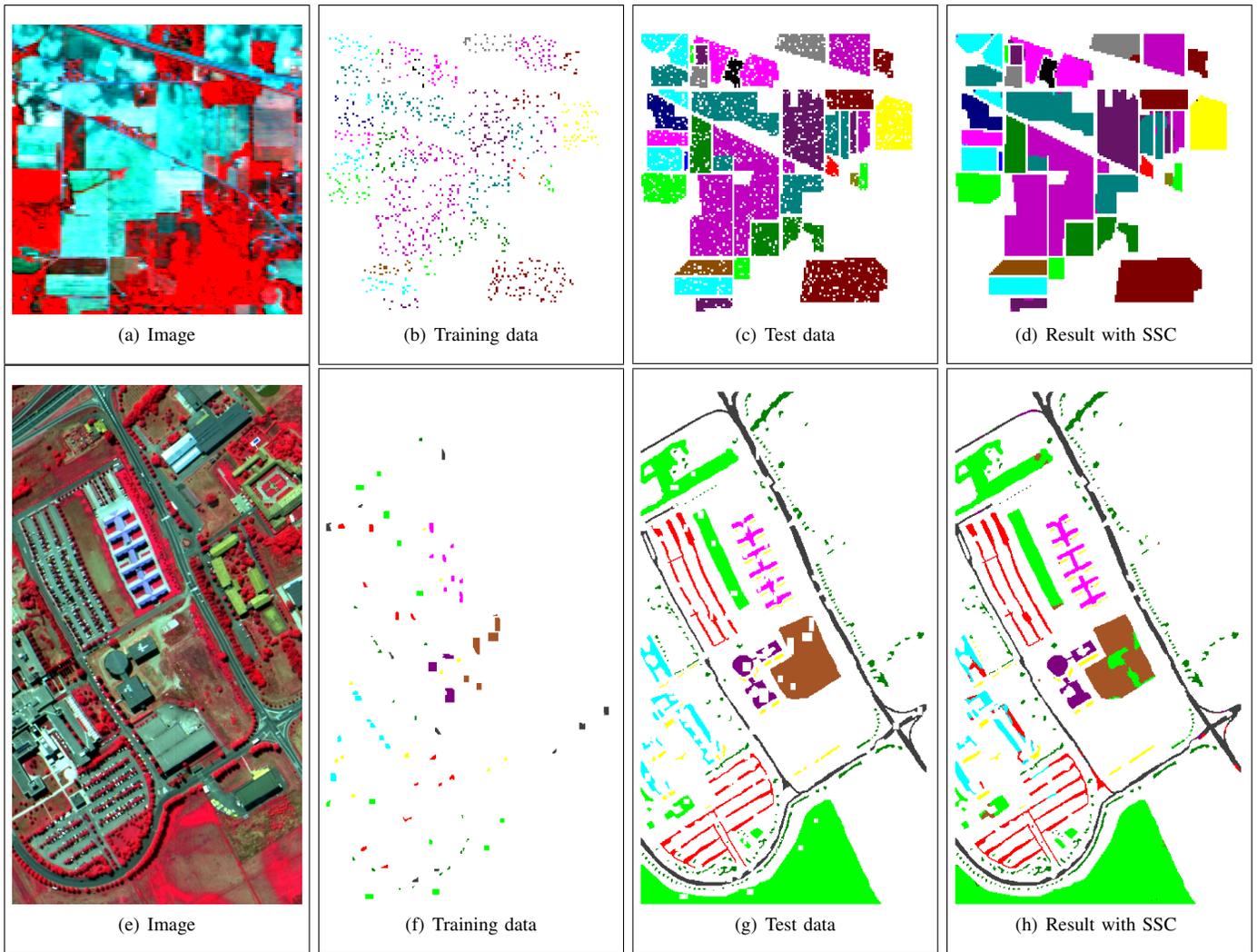
Fig. 4: Image, training and test data of the data sets INDIAN PINES (upper row) and UNIVERSITY OF PAVIA (bottom row). The classification result of the SSC approach is shown in the right column.

for hierarchical sparse coding," *The Journal of Machine Learning Research*, vol. 12, pp. 2297–2334, 2011.

[27] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.

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

[29] Y. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," in *Conference Record of The Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, 1993, pp. 40–44 vol.1.

[30] U. Srinivas, Y. Chen, V. Monga, N. Nasrabadi, and T. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 505–509, 2013.

[31] J. Chua, I. Givoni, R. Adams, and B. Frey, "Learning structural element patch models with hierarchical palettes," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2012, pp. 2416–2423.

[32] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods,"

IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 11, pp. 2274–2282, 2012.

[33] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002.